2019

Association of Adaptive Early Phase Study Design and Late Phase Study Results in Oncology

Donna Elise Levy
Walden University

Walden University

College of Health Sciences

This is to certify that the doctoral dissertation by

Donna Elise Levy

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. Fred Tabung, Committee Chairperson, Public Health Faculty
Dr. Srikanta Banerjee, Committee Member, Public Health Faculty
Dr. Jagdish Khubchandani, University Reviewer, Public Health Faculty

Chief Academic Officer
Eric Riedel, Ph.D.

Walden University
2019

Abstract

Association of Adaptive Early Phase Study Design and Late Phase Study Results in Oncology

by

Donna Elise Levy

MA, Boston University, 2007

MS, University of Arkansas, 2000

MS, University of Arkansas, 1998

BS, Oklahoma State University, 1993

Dissertation Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

Public Health – Epidemiology

Walden University

May 2019

Abstract

This quantitative study assessed the association of the design methods used for early phase oncology studies (adaptive versus traditional) and the outcome of late stage clinical trials. Differences by cancer type and by drug classification were also assessed. The theoretical and conceptual frameworks used were the general systems theory and the design and evaluation of complex interventions, respectively. Units of analysis were individual oncology studies in the ClinicalTrials.gov database and Bayesian logistic modeling was applied on a random sample of 381 studies initiated after November 1999 to December 2016. When assessing study design and outcome, there were lower odds of a positive outcome when adaptive methods were used though this association was not statistically significant (OR [95% highest posterior density (HPD)]:0.66 [0.20, 1.21]). Among the different drug types, using adaptive compared to traditional methods was associated with significantly higher odds of a positive outcome for taxanes, OR: 2.75, 95% HPD: 1.01, 5.16) and other, OR: 3.23, 95% HPD: 1.58, 5.46) but no association among studies of monoclonal antibodies or protein kinase inhibitors. Also, there were no significant associations between early phase study design and outcome in late phase studies by cancer type (lung, breast, other). Further research should be conducted using all completed oncology clinical trials in the database to more precisely determine the relationship between adaptive study design in early phase oncology studies and outcomes in late stage studies. Social change can occur through increased uptake of adaptive design methods, which may lead to more efficacious cancer treatment options.

Association of Adaptive Early Phase Study Design and Late Phase Study Results in Oncology

by

Donna Elise Levy

MA, Boston University, Biostatistics, 2007

MS, University of Arkansas, Statistics, 2000

MS, University of Arkansas, Applied Mathematics, 1998

BS, Oklahoma State University, 1993

Dissertation Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

Public Health – Epidemiology

Walden University

June 2019

Dedication

This dissertation is dedicated to my dad, J. Douglas Levy who frequently asked me, to my frustration, ‘when I was going to be done’, until he no longer remembered to ask. I wish he would ask me that same question now. He is my inspiration for the love of numbers and mathematics. He supported me through my athletics in high school and university by frequently driving out to isolated locations and waiting patiently so that I could get my workout completed. His support of my academics and athletics continued through my adulthood, including shaking his head when he did not understand why I had pushed so hard in some race or another. I hope you will allow me to support you in your golden years as much as you supported me.

I would also like to dedicate and thank my other half, Daniel Quinn, for his support over the years to complete this work, including his repeated review of my content, discussions on the topic, database administration and SAS programming guidance. Being able to study, then having a home cooked meal ready to allow for continued studying after dinner was also appreciated. This journey truly was a team effort with a lot of time and sacrifice from both of us.

Acknowledgments

I would like to express my deepest gratitude and appreciation for my committee chair, Dr. Fred Tabung. His knowledge, guidance, editing and content suggestions were truly appreciated through the initial discussions of the topic as well as the dissertation process. I would also like to thank Dr. Srikanta Banerjee for his encouragement and enthusiasm during the coursework phase of this degree, as well as during the dissertation process. I am sure that every student that you have had appreciates your wealth of knowledge as well as your vigor to continue learning.

This dissertation would not have been possible without the insight of the NIH to create a clinical research registry, as well as the expanded requirement need to include study results. This information and free access is critical for patients and practitioners to learn about clinical trials treatment, related to the disease of interest. The database also allows for information dissemination of treatment for researchers as well. Sharing of information is critical in clinical research to allow for a continued path for effective treatment development. I also thank the cancer patients that participated in the clinical trials included in the database and this research. While I hope the treatments helped their life, their selfless and altruistic behavior will help patients in the future.

Thanks to the researchers that willingly shared their data and analysis for this research, specifically Dr. Aaron Kesselheim (Bothwell, Avorn, Khan, & Kesselheim, 2018) and Dr. I Hatfield (Hatfield, Allison, Flight, Julious, & Dimairo, 2016) of which included their adaptive search terms and trial information on a public website. With their

desire for collaboration as well as the importance for continuing to move the field forward, this type of collaboration is needed in all areas of clinical research.

Also thanks to colleagues Maggi Beckstrand, Eva Miller, Bingyan Wu and Zejiang Yang for their input, guidance and discussions related to the analyses conducted through the dissertation process. Their statistical expertise and support starting from the development of this study through the final analysis was much appreciated. I would hope that all researchers have colleagues that provide undo support like I have had throughout my entire career. Thanks also to Dan Gower and Daniel Quinn for their database administration, database programming, data extraction assistance and, technical expertise. Without their help, this study and analysis would not have been possible.

Finally I would like to say thank you to the Walden University faculty and staff including Jagdish Khubchandani for his thorough University Research Review; the library and writing staff, including Dr. Joe Gredler with his APA expertise and patience.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction to the Study

Cancer rates are increasing in the United States and across the globe (Centers for Disease Control and Prevention, 2017). While cancer has been reported as a single disease with a single treatment, cancer can originate from a variety of locations in the body (National Cancer Institute, 2015), be diagnosed at multiple stages of development and have varying genetic markers that can impact treatment response, leading to the complexity in treatment development. However, treatment development and approval have stagnated over the years, while costs have increased and the success rates of clinical trials have reduced (Berry, 2012; Christopher S. Coffey, 2017; Prasad & Mailankody, 2017). Oncology treatments have poor regulatory approval rates with failure rates as high as 66% in results reported from 2003 to 2010 (Berry, 2011). In comparison to other diseases, oncology studies have been reported to have the lowest likelihood of regulatory approval (Biotechnology Innovation Organization [BIO], 2016). With more than 30% of the treatments in development, approval for oncology drugs progressing through the clinical trial process was reported to be only 5.1%. The Food and Drug Administration has also been trying to encourage innovation to improve treatment approval through the Critical Path Initiative (U.S. Food and Drug Administration, 2004). Improved and innovative design methods adopted in clinical trials should lead to improved and increased treatment options for patients with cancer.

In this chapter, I provide background information with respect to the burden of cancer and the need for innovative methods given the limited success in cancer clinical trials with increasing costs. I also present an overview of the purpose of the study, the

study design, research questions and hypotheses. The components of the study are constructed in the theoretical foundation and conceptual framework. The assumptions of the study, analysis methods, as well as limitations, are also discussed.

Background

In comparison to other modalities, cancer clinical trials are the most frequent in clinical research, however, the success rate is quite poor (Biotechnology Innovation Organization [BIO], 2016). While the costs of oncology research continue to increase, researchers have noted that the oncology clinical trial failure rate is 66% for clinical trials reported from 2003 to 2010 (Berry, 2012; S.-C. Chow & Chang, 2008). Traditional methods developed in the 1940s continue to be used in the majority of early phase oncology studies (Hansen, Graham, Pond, & Siu, 2014). However, the efficacy associated with these traditional methods need to be assessed in order to improve treatment options for oncology patients. In addition, the traditional methods continue to be utilized even when the underlying monotonic dose-toxicity assumption is not held (Hansen et al., 2014).The limitations of the traditional methods need to be addressed as patients are exposed to sub-therapeutic doses due to the conservative methods (Butterfield, Disis, Khleif, Balwit, & Marincola, 2010; Hansen et al., 2014; Le Tourneau, 2009). The traditional methods also identify the appropriate dose level only 30% of the time (Rogatko et al., 2007). With these limitations on traditional design options, innovative methods need to be developed, adopted and assessed.

Barker et al. (2009) found that adaptive methods that use predictive models of therapeutic responses can be utilized to increase the speed of drug development.

Specifically, the I-SPY2 (Investigation of Serial Studies to Predict Your Therapeutic Response With Imaging and moLecular Analysis 2) study includes innovative, adaptive design methods and is supported by academic and regulatory bodies, has been prolific with respect to drug development in breast cancer. The study includes continuous enrollment and rolling treatment assessments, both of which can be aspects of adaptive design and integrates learning across and within included treatment components, but does not compare study design methods and outcomes. Researchers noted the need for innovative methods and improved study design in clinical trials to increase safety, efficacy and quality of drug development leading to improved treatment patient profiles and treatment options for patients. In order to improve study outcomes, innovative methods need to be better understood and utilized (Parekh et al., 2015).

Berry (2012) noted that the costs associated with oncology drug development have been increasing over the years without the corresponding trial success, thus the need for improved and innovative studies. With the increasing cost of treatment development, more efficient and innovative study design methods need to be developed, assessed and utilized (Berry, 2012). However, barriers related to the ease of the innovative methods adoption have been identified regardless of the methods assets. The researchers noted the current barriers related to infrastructure, such as access to software, method understanding, knowledge and ease of use have led to a lower rate of adaptive design method adoption (Kairalla, Coffey, Thomann, & Muller, 2012; J. Quinlan, Gaydos, Maca, & Krams, 2010). More studies need to be conducted comparing operating

characteristics of traditional versus adaptive methods as well as adaptive design case studies to increase the knowledge base.

Berry (2011) noted that both traditional and adaptive methods could be utilized in different settings using the strengths of each of the design methods. However, researchers need to understand the strengths and the operational characteristics of each method so they can be appropriately utilized when developing a clinical trial. Hatfield et al. (2016) assessed adaptive studies reported in CT.gov (2000 to 2014) and found that there has been a threefold increase in adaptive method use particularly in oncology of the study period. However, studies including adaptive design methods are approximately 25% (143 of 573 studies) of the total number of clinical trials of phase II or higher. Recently researchers also used the ClinicalTrials.gov (CT.gov) database and found that approximately 20% of oncology clinical trials of phase II or higher used adaptive methods (Bothwell et al., 2018). While Bothwell et al. (2018) noted that 49% of the adaptive studies were deemed effective, neither Hatfield et al. (2016) nor Bothwell et al. (2018) compared the adaptive versus traditional designed clinical trial results or their downstream impact. More studies need to be conducted comparing operating characteristics of traditional versus adaptive studies.

Problem Statement

Cancer is the second leading cause of death for men and women in the United States (Centers for Disease Control and Prevention, 2017). The most commonly diagnosed cancers in 2017 were projected to be breast, lung, bronchus, prostate and colon cancer (National Cancer Institute, 2017). More than 1.6 million cases of cancer were

estimated to be diagnosed in the United States and over a half million individuals were reported to have died from the disease in 2016 (National Cancer Institute, 2017). With the increasing demands of cancer, treatment and care of cancer patients in the United States, effective treatments need to be developed through clinical trials with therapeutic dose levels and targeted populations.

Using carefully planned safety, efficacy assessments, targeted cancer type and appropriate biomarker risk factors, effective treatments can be identified using efficient accrual and study design methods (Barker et al., 2009). Oncology treatment characteristics, specifically drug classifications including related biomarkers and tumor types, must be considered in study development (Barabási, Gulbahce, & Loscalzo, 2011; U. S. Food and Drug Administration, n.d.-b; Siddiqui & Rajkumar, 2012). Cancers are diverse in their location of development; treatment paths and biomarkers aid in learning about the target population and possible future standard of care (Ardies, 2014; Barker et al., 2009; Berry, 2012). Studies using biomarkers have also been shown to have increased likelihood of approval in comparison to studies without (Biotechnology Innovation Organization [BIO], 2016). Strengthening study design components should be addressed for improved clinical research. Unreliable surrogate endpoints (Fleming, 1996), limitations in disease response measurement techniques (Weber, 2009), inadequate animal models (Ananthakrishnan et al., 2017; Butterfield et al., 2010), bias introduced with historical comparators as well as limited clinical biomarkers (Butterfield et al., 2010; U. S. Food and Drug Administration, 2017) may lead to false positive or false negative results in early phase studies. Innovative study design methods that overcome some of

these limitations also need to be developed, assessed and adopted in early phase oncology studies.

Within the field of oncology clinical trials, while early phase studies (Phase 0, 1 and 2) show promising results, more than 66% of late stage (Phase 3) studies reported from 2003 to 2010 do not lead to statistical significance or positive results for the patient population (Berry, 2012). Additional research was conducted assessing the success of oncology studies from 2006 to 2015 and similar results were reported (Biotechnology Innovation Organization [BIO], 2016). The low rate of success in clinical research is a concern that needs to be addressed.

Often traditional or fixed design methods are being used in early phase studies, as the methods have been historically accepted. The most common design method used in early phase clinical trials is the 3+3 design (Hansen et al., 2014). The 3+3 design enrolls patients by group (usually groups of three) with respect to protocol defined treatment dose level. In the initial set of patients, the individuals are assessed for dose limiting toxicities (DLT) in the DLT assessment period (e.g. one cycle of treatment). If fewer than a prespecified number of patients experience a DLT, the next set of patients are enrolled at the next highest dose. If the toxicity rate is higher than the planned rate, then additional patients may be enrolled at the same dose level or the study is stopped and the maximum tolerated dose (MTD) is the previous dose level. Unfortunately, initial doses in the 3+3 design are often conservative as they usually start at a subclinical dose (Hansen et al., 2014; Kairalla et al., 2012). Researchers have also found that these traditional design methods are often used in clinical trials with inappropriate drug classifications such as

molecularly targeted agents (MTAs). In traditional methods, there is an assumed dose-toxicity monotonic relationship; as the dose increases, the probability of toxicity also increases (Ananthakrishnan et al., 2017). For MTAs and other recent oncology drugs that have been recently developed, the monotonic dose-toxicity relationship is not satisfied. In MTAs, toxicities are related to the accumulation of the drug, thus the assumed treatment-dose relationship within a specific cycle is not satisfied (Hansen et al., 2014; Kairalla et al., 2012). However, the traditional methods continue to be utilized, in spite of the underlying assumptions of the design method not being met.

Even with the research indicating the ineffectiveness of traditional methods, the adoption of innovative methods has been sporadic due to limited knowledge, time and infrastructure needs not consistently being met (Christopher S. Coffey, 2017; Hansen et al., 2014; Hatfield et al., 2016; Kairalla et al., 2012). Regulatory bodies including the Food and Drug Administration (FDA), Committee for Medicinal Products for Human Use (CHMP) to name a few, have noted the stagnation of innovation, a reduction in effective treatment identification, as well as challenges and opportunities in diseases such as cancer (Committee for Medicinal Products for Human Use [CHMP], 2006; U. S. Food and Drug Administration, 2006, n.d.-a). Innovative methods such as adaptive study designs may improve treatment development in oncology clinical research. Adaptive design clinical trials are defined to be studies with preplanned opportunities for modifications for one or more aspects included in the study design (Committee for Medicinal Products for Human Use [CHMP], 2006; U. S. Food and Drug Administration, 2010). Design elements that can be modified within the adaptive plan include the sample

size, treatment assignment, population, dose level, among other study characteristics (Bornkamp et al., 2007). Any initial study characteristic can be adapted during a study, as long as the adaption is prespecified. Oncology clinical trial design is an area in need of innovative method development, assessment and adoption, including adaptive methods.

Purpose of the Study

Although adaptive methods have been available since the 1970s and with appropriate computational support and knowledge since the 1990s (A. R. Brown et al., 2016), researchers have been reluctant to utilize the methods. However, uptake of adaptive methods has been assertively accepted at some research institutions likely due to awareness, training, education, feasibility, expertise, computational and programming availability (S. C. Chow, Corey, & Lin, 2012; J. A. Quinlan & Krams, 2006; Viele & McGlothlin, 2017). Appropriate communication with the study team to increase comfort understanding and communication with regulatory bodies is also critical in the study development process (Kairalla et al., 2012; Viele & McGlothlin, 2017). Awareness, training and expertise need to be improved to allow for increased innovative methods.

In this quantitative study, I investigated the potential association of early phase study design research outcomes that utilize traditional versus adaptive methods on late stage results in oncology clinical trials. By comparing traditional versus adaptive methods and the downstream clinical trial outcome, the benefits related to late stage study success may be determined. Success in late stage studies may lead to improved patient treatment options. As drug development and approval are based on cancer type, study phase and treatment classifications, these potential effect modifiers were included in the analysis

model (Barabási et al., 2011; U. S. Food and Drug Administration, n.d.-b; Siddiqui & Rajkumar, 2012), to determine whether the association between use of adaptive versus traditional methods (e.g.. 3+3) in early phase studies and the results outcome of the late stage study, was found to be different in subgroups defined according to levels of these factors.

Analysis Model

The variables included in the analysis model are defined in Table 1. The analysis model is: Late phase study results (favorable/unfavorable) = early phase design (adaptive/traditional) + early phase + experimental treatment classification + type of cancer + drug classification + duration between early and late phase studies (months) + study funding

Table 1

Analysis Model Variables, Description and Possible Values

Variable	Definition	CT.gov format	Values
Dependent variable			
Outcome of late-phase study	Late-stage study results. Results are based on the original study operation characteristics and statistical significance. The statistical outcome of the late-phase study (favorable = experimental treatment statistically better than comparison; equivalent = experiment and comparison are statistically equivalent; nonfavorable = experimental treatment is found to be better than the comparison treatment). For studies that are not driven by statistical significance, clinical relevance (effect size) will determine the outcome classification.	Free text, categorized for variable values.	Endpoint: favorable, equivalent, unfavorable
Independent variables			
Design classification	Adaptive studies are defined to be clinical study designs that use accumulated information or data of the study to modify aspects of the study as it continues (Christopher S Coffey et al., 2012). The adaptations in these designs are predefined. Traditional studies will be defined to be nonadaptive studies.	Free text	Adaptive, traditional
Early phase	Phase of early-stage study	Same as value	Categorical: 0, 1a, 1b, 1, 2, 2a, 2b
Experimental treatment classification	Interventional treatment classification	Categorical	Drug Device Biological/vaccine, Procedure/surgery Radiation Combination Other
Cancer type	The cancer site or type that is being researched in the early- and late-stage study.	Same as reported value	Breast, lung, pancreatic, bone etc.

(continued)

Table 1 Continued

Variable	Definition	CT.gov format	Values
Experimental drug classification	WHO drug classification	Categorical variable	EGFR or VEGF inhibitors, monoclonal antibody, proteasome inhibitor, immunotherapy, etc.
Duration between studies (months)	Time between the end of the early-stage study and the beginning of the late-stage study	Dates; will take difference for duration	Time in months
Study funding source	Who is sponsoring the early-stage study?	Reported funding (grant, other)	Private or public funding
Sensitivity analysis variables			
Type of endpoint	Type of endpoint used. Any endpoint that is not overall survival such as progression free survival or objective response will be considered a surrogate.	Free text	Surrogate, clinical, or both
Sample size	The number of patients planned to be enrolled in the early-phase studies.	Numeric	Numeric
Biomarker	Did the study include a biomarker as an endpoint in the study or not?	Categorical	Binomial (yes, no)
Adaptive classification	For those studies that are adaptive, are they are classified as well-understood or less well-understood designs as defined by the FDA (2010).	Categorical	Well-understood, less well-understood, not adaptive

Research Questions and Hypotheses

RQ1: To what extent is there an association between design methods used for early phase oncology studies (adaptive versus traditional) and the outcome of late stage clinical trials?

H_o1: There is no association between early phase oncology studies (adaptive versus traditional) and the late stage clinical trial outcomes.

H_a1: There is an association between early phase oncology studies (adaptive versus traditional) and the late stage clinical trial outcomes.

RQ2: In specific cancer types (e.g., breast cancer, lung cancer etc.), what is the association between the design methods used for early phase oncology studies (adaptive versus traditional) and the outcome of the late stage clinical trials?

H_o2: The association between early phase oncology design methods (adaptive versus traditional) and the outcome of the late stage clinical trials does not differ between specific cancer types.

H_a2: The association between early phase oncology design methods (adaptive versus traditional) and the outcome of the late stage clinical trials differs between specific cancer types.

RQ3: How does the treatment classification modify the relationship between the design methods used for early phase oncology studies (adaptive versus traditional) and the outcome of the late stage clinical trials?

H_o3: The association between early phase oncology design methods (adaptive versus traditional) and the outcome of the late stage clinical trials does not differ between specific treatment classification (e.g. surgical, adjuvant, radiation, etc.).

H_a3: The association between early phase oncology design methods (adaptive versus traditional) and the outcome of the late stage clinical trials does differ between specific treatment classification (e.g. surgical, adjuvant, radiation, etc.).

Theoretical Framework

The general systems theory developed by Kenneth Boulding in 1956 is the study of systems and how the interrelated and interdependent parts interact with each other (Boulding, 1956). The underlying assumption of the theory is that as one component of the system changes, this change will affect other components of the system. In this study, I assessed the association between early phase design methods, specifically traditional versus adaptive and how the design method is associated with the late stage study outcome (favorable, equivalent or nonfavorable). The underlying hypothesis was that when adaptive methods are used, increased and high-quality information is gained and used in the design of late stage studies, leading to improved results in late stage studies. Adaptive methods may not be as effective in every type of cancer, treatment classification, or experimental drug classification, but the impact of the study characteristics will be assessed when adaptive methods are used versus not. Within systems theory, changing or altering one component can impact another (Boulding, 1956). The change of one study characteristic and the impact in other components is evident in clinical trial development.

For example, changing one classification, say target population, will likely impact the type of treatment as well as drug classification, as cancer is a diverse number of diseases with a variety of treatment paths due to a variety of cell proliferation paths (Ardies, 2014; Lonial & Nooka, 2016; National Cancer Institute, 2015). In addition, using the same treatment in different target populations can impact the outcome of the study, as not all treatments are effective in every population. The hypothesis of this study

was that the component design method (adaptive or traditional) is associated with the downstream results of the later stage study. The hypothesis is based on the idea that using the innovative design methods, the late stage outcomes may be improved due to the quality and increased quantity of gained information.

Conceptual Framework

The Framework for Design and Evaluation of Complex Interventions (DECI) provides guidance on the integrated evaluation of complex interventions (Campbell et al., 2000). The framework was developed to assess complex research interventions with various interconnected components. The information gained within a study as well as the study results are integrated within and across every phase in an iterative fashion (Campbell et al., 2000). Due to the complexity of treatment and interventions in clinical trials, learning across and within studies is critical to continue learning and improving clinical trial outcome as well as assessing unexplored endpoints (Geifman & Butte, 2016). Further learning in and across studies with positive as well as negative findings may aid in advancing the field (Butterfield et al., 2010).

Researchers have noted the need for well-designed and implemented studies as well as the evaluation of the studies through meta-analysis and long-term programs, utilizing the DECI methods of systematic review and iterative information gain (Grol, 2001). Due to the sequential nature of oncology clinical research, an iterative approach should be consistently applied and interventions re-examined as needed information is collected (Campbell et al., 2000; Geifman & Butte, 2016). Thus, the early phase studies are not just providing guidance for late stage studies but are also providing input in

combination with other research used for other ongoing and future early phase studies that use the treatment or directed toward the indication. Using adaptive methods may lead to better designed studies and improved target population information that can be used for integration into other studies. In addition, systematic reviews of clinical trials and health related data are critical for improving methods as well as treatment for the targeted populations (Griffiths, Lindenmeyer, Powell, Lowe, & Thorogood, 2006; Grol, 2001; Kroeze, Werkman, & Brug, 2006; Steinert et al., 2006). This research included oncology studies randomly identified in the ClinicalTrials.gov database. The focus of this research was to quantitatively assess adaptive methods versus traditional methods used in oncology clinical trials. I used early clinical study research, which is used to develop late stage clinical studies. The interrelationship of the phase of treatment is a key component of DECI.

Nature of the Study

In this study, I used a quantitative methods approach. Because I assessed the association of early phase design methods and the downstream influence on late stage clinical trials results in oncology numerically, a quantitative assessment of the outcome was appropriate. I extracted data from the National Institutes of Health (NIH) Clinical Trials registry and results database (National Institute of Health, 2017a). If not provided, the experimental treatment was classified as needed using World Health Organization (WHO) drug classification (World Health Organization Collaboration Centre, n.d.).

The phase of the study, the date associated with the results (to ensure sequential studies) and the design methods (adaptive versus not) were utilized in the model to determine if there is a relationship between early stage design methods and late stage results in oncology studies. Experimental treatment classification, drug classification and type of cancer were also included in the model. While there are strengths of traditional and adaptive methods, researchers need to be able to utilize the appropriate methods, such as drug classification and type of cancer as well as understand the operating characteristics of the study methods (Berry, 2011). This study may aid in identifying oncology populations and treatment classifications where adaptive methods are more effective leading to improved study outcomes in late stage studies. Additional sensitivity analyses were conducted including whether the study included a surrogate endpoint or not, planned sample size as well as adaptive method classification.

The data used for this analysis was extracted from the National Institutes of Health Clinical Trials registry and results database (National Institute of Health, 2018). The database was established in 2008 where study characteristics and results are required for specific clinical trials, including oncology studies. Data in the database includes aggregate participant information, baseline characteristics, primary outcome measures, statistical analyses as well as adverse event information. Individual studies were the units of analysis. Participant level information is not available in this database, thus was not used as the unit of analysis for this research study. WHO Drug classification as well as study classification (adaptive versus traditional) was added to the dataset based on treatment and study design information provided in the database or using supplemental

publications. As researcher contact was included in the database, study characteristics or results that were unclear were clarified through contacting the study researchers.

Operational Definitions of Terms

Adaptive or traditional studies: Clinical study designs that used accumulated information or data of the study to modify aspects of the study as it continues (Christopher S Coffey et al., 2012). The adjustments are predefined. This definition is consistent with the Pharmaceutical Research and Manufacturers of America (PhRMA) Working Group (Gallo et al., 2006). Traditional studies were studies that were not classified as adaptive.

Cancer type: The sub-population of cancer or cancers treated in the reported study.

Duration between studies: The difference between the date reported in the ClinicalTrial.gov database for the late stage and early stage study. Only positive time differences (and their associated studies) were included in the analysis indicating that the early stage study occurred before the late stage study.

Experimental drug classification: Based on the World Health Organization (WHO) Drug Dictionary which was based on the Anatomical-Therapeutic-Chemical (ATC) classification system (World Health Organization Collaboration Centre, n.d.). For example, anti-cancer drug bevacizumab was classified as “Monoclonal antibodies (L01XC)” within “Other antineoplastic agents (L01X)” using the WHO drug classification dictionary.

Experimental treatment classification: Based on the study treatment defined in the study such as surgery and radiation to name a few.

Late stage results: Definitions of late stage study outcome followed a similar definition utilized by Rasmussen et al. (2009). Late stage studies that were considered favorable were those reported in the clinicaltrials.gov database that were statistically significant based on original study criteria (e.g. $p < 0.05$) in favor of the experimental treatment (Rasmussen, Lee, & Bero, 2009). For studies that were not driven by statistical significance, clinical relevance (effect size) was used for the outcome classification. Unfavorable studies were defined to be studies where the results were statistically significant in favor of the comparator treatment or did not meet the definition of favorable. For this study, equivalent studies were defined to be studies where the results were deemed equivalent (favorable).

Less well-understood adaptive methods: methods that are not frequently utilized thus there are less experiences with the methods (U. S. Food and Drug Administration, 2010).

Phase of the study: The clinical research phase reported by the study researchers. Early phase studies were defined to be phase 0, I, Ia, Ib, II or any combination (phase 2 or less). Late phase studies were defined to be phase III, IIIa, IIIb or phase IV studies (phase 3 or higher).

Study funding source: The primary source of funding of the research. An example of private funding would be a pharmaceutical company. An example of public funding

would be the National Cancer Institute. A study could have multiple funding sources and were included in the model to reflect the multiple sources.

Type of endpoint: Sensitivity analyses included the type of endpoint of the early phase study. Overall survival is considered the gold standard in oncology studies. Any other endpoint was considered a surrogate for this analysis. If the endpoint was a measure of the disease or a laboratory abnormality, the endpoint was classified as clinical. Endpoints were classified as clinical, surrogate or both. Similar definitions of endpoints were used by the researchers, Bothwell et al. (2018).

Well-understood adaptive designs: These designs are frequently used adaptive methods and include group sequential methods with unblinded interim data review and controlled Type I error.

Assumptions

For this study, I assumed that the clinicaltrials.gov database was representative of all oncology clinical trials. Applicable studies that were initiated after January 1, 2000 should be registered within the database according to federal guidelines for United States enrolling studies (National Institute of Health, 2017b). All phases of studies with one or more sites in the United States and all intervention types should be included in the database. Applicable studies include FDA-regulated drugs, biological products or devices that meet one of the following conditions: trial conducted under an FDA investigational new drug application or investigational device exemption or the trial involves a drug, biologic or device that is manufactured in the United States or its territories and is exported for research. Registering within a public database before patients begin

enrolling is also required by most medical journals in order for articles to be accepted for publication (International Committee of Medical Journal Editors, 2017). In addition, National Institutes of Health (NIH) supported trials are also encouraged to register. The Declaration of Helsinki also states that all research studies involving human subjects must be registered and researchers have the responsibility to make the research publicly available (National Institute of Health, 2015b). As a result of these regulations, requirements and guidance, the clinicaltrials.gov database registry and results should be representative of oncology clinical research studies.

This study also assumed that the results database was up to date with results that were included in the analysis and that there was no systematic pattern (Missing at random, missing completely at random) of delayed reporting and the study outcome. Based on the requirements for the Clinical Trials database, most studies have one year from primary endpoint completion (National Institute of Health, 2017a) to have their results entered in the database, which should result in a reduction in delayed results reporting.

Scope and Delimitations

The scope of this study assessed the association with early phase oncology design methods and the outcome of late stage outcome results. This study was delimited to include oncology studies and human interventions only. In addition, this research only included late stage studies that have results in the Clinical Trials database. The research was delimited to include phase, experimental treatment classification, cancer type, experimental drug classification, duration between studies and study funding source. Late

stage studies were also not included as independent variables in the model even though they likely provide information for target population and effective treatment identification for future studies.

Limitations

For any research conducted, limitations of the research may be identified. A limitation of this research was the possible accuracy of the database. Researchers have noted that errors have been found in the Clinical Trials database (Hartung et al., 2014). However systematic errors have not been identified related to the type of study such as adaptive designs versus traditional studies. Early database entry studies likely have a higher error rate (Hartung et al., 2014). As such 10% of the studies' results included in the analysis were confirmed through publication review as well as reaching out to the researchers to confirm the data reported in the clinical trials database.

While a programmatic identification of adaptive studies was utilized, with limited structure related to adaptive classification and the frequency of free text utilization within the Clinical Trials database, likely not all adaptive studies were identified. In addition, with the ambiguity and changing classification of adaptive design methods, this likely lead to a limitation of systematically identifying adaptive studies. Other researchers have also identified the database structure and adaptive design definition a limitation in utilization of the Clinical Trials database (Bothwell et al., 2018; Hatfield et al., 2016). Once again, 10% of studies included in this analysis were manually reviewed and design classifications were compared to publication results or confirmed with Clinical Trials study researchers.

Significance

Researchers have noted that drug development and approval has been stagnating even though the cost of development has been increasing (Barker et al., 2009; Berry, 2011, 2012; U. S. Food and Drug Administration, n.d.-a). The results of this study may aid in the understanding of the influence of adaptive methods and late stage clinical trial outcomes as well as the influence of treatment populations and drug classification. While adaptive design methods in clinical trials have increased in use threefold since 2001, the methods are used in less than 30% of studies (A. R. Brown et al., 2016; Hatfield et al., 2016). This research project may inform researchers on the importance of innovative study design methods due to the current low success rate of studies using traditional methods (Hansen et al., 2014; Rogatko et al., 2007). This study may increase knowledge related to adaptive methods and scenarios where the methods seem to provide improved results, reducing the barriers related to method utilization (Hatfield et al., 2016; Kairalla et al., 2012; Rogatko et al., 2007). The results of this study may aid in reducing the identified barriers with respect to the acceptance and utilization of adaptive design methods in oncology clinical trials. Inference from this study may inform researchers on the need for infrastructure, education and knowledge with respect to adaptive study designs leading to increased usage of the methods. Ultimately, this study may contribute to the identification of more effective treatments, as well as earlier identification of ineffective treatments, thus improving oncology patient care and outcome. The efficacy of adaptive methods are often shown in simulations, but have not been assessed in real clinical trials (Ananthakrishnan et al., 2017). Advances in oncology research are needed

and innovative methods such as adaptive design may improve treatment options for cancer patients.

Summary

With the increasing rate of cancer diagnosis, effective treatments need to be identified and developed. Further, patients need to be limited to the exposure of ineffective or sub-therapeutic treatment doses. Also with the increasing cost of drug development and stagnate results, innovative designs are needed to be assessed and adopted. While traditional methods are most frequently utilized in early phase studies, adaptive methods should be adopted as simulations studies indicate that these methods are more efficient at identifying effective and ineffective treatments. This study assessed the association of early phase design methods and late stage outcome results in oncology clinical trials.

Chapter 2: Literature Review

Often traditional or fixed methods are being used in early phase studies, as the methods have been historically accepted. The most common design method used in early phase clinical trials is the 3+3 design with utilization rates as high as 98% (Hansen et al., 2014; Rogatko et al., 2007). The traditional 3+3 design was introduced in the 1940s as a method to systematically escalate dose and monitor treatment safety (Bauer & Einfalt, 2006; Dimairo, Boote, Julious, Nicholl, & Todd, 2015; Hatfield et al., 2016). The 3+3 design enrolls patients by group (usually in groups of three) and protocol defined treatment dose level. At each treatment level, dose limiting toxicities (DLT) are assessed and based on the rate of toxicities, the dose level is increased, additional patients are accrued at the same dose, or the study ends. Treatment is assigned based on prespecified rules and dose escalation can only occur if there are fewer DLTs than planned or expected (Le Tourneau, 2009).

Initial doses in the 3+3 design tend to be conservative due to subclinical dose levels (Hansen et al., 2014; Kairalla et al., 2012). In addition, the traditional methods have been shown to identify the appropriate dose or treatment approximately 30% of the time (Simon et al., 1997). In traditional methods, there is an assumed dose-toxicity monotonic relationship; as the dose increases, the probability of toxicity also increases. However, the assumption may not always be met in newer treatments as toxicity may be related to dose accumulation rather than dose intensity (Ananthakrishnan et al., 2017; Butterfield et al., 2010; Kairalla et al., 2012).

Traditional methods continue to be used in oncology studies even though newer, more innovative, efficient and effective design methods are available. Innovative methods such as adaptive methods use information learned from within the studies in order to adjust the underlying assumptions within the study itself. A guidance document for adaptive trials was released in 2010 by the Center for Drug Evaluation and Research (CDER), the Center for Biologics Evaluation and Research (CBER) and the Food and Drug Administration (FDA). As a result of these initiatives, adaptive approaches have been increasingly utilized. Through grants and educational opportunities, researchers have attempted to overcome the slow acceptance and utilization of innovative methods. (Berry, 2012).

This chapter includes a discussion of the issues related to clinical research development in oncology and the decision making related to study design methods. Many factors impact the development of cancer treatment, including the type of cancer, subtype and related risk factors. All of these components are considered when developing a clinical trial and addressed during study design selection and development. Clinical trials go through prespecified phases of development, of which the earlier studies lead to the underlying assumptions and operating characteristics that are used in the late stage studies. While there may be advantages to traditional and adaptive methods, disadvantages of both design methods will be discussed. With the adoption of innovative design methods, there are barriers that also need to be identified, addressed and overcome with reasonable solutions. With the dismal results of oncology clinical trials with a failure rate as high as 66% for studies reported from 2003 to 2010 (Berry, 2011), all aspects of

research should be considered for improvement, including composite and innovative endpoints. The aim of conducting this research was to find the potential influence of early phase study design research outcomes that utilize traditional versus adaptive methods on late stage results in oncology clinical trials.

Literature Search Strategy

In this literature review, I will provide a summary of current adaptive design research as well as the state of clinical research in oncology. Oncology related summaries specifically treatment, risk factors and trial results were included in the review. Search terms included: *adaptive design oncology trial success, clinical trial phase of treatment, clinical trial, oncology risk factors, oncology, surrogate endpoints, 3+3 and adaptive clinical trials*. Specific design methods were also used in the search strategy such as Continual Reassessment Method (CRM), Sample Size Reassessment (SSR) and accelerated 3+3. Leaders in the field of adaptive methods including Peter Bauer, Scott and Don Berry, Christopher Coffey and Y.H. Joshua Chen were also included in my literature search.

I searched for references within all Walden databases including ProQuest and PubMed databases as well as Google Scholar. Peer reviewed technical articles published after January 1, 2009, were included in my search, unless the work was considered critical and referenced frequently by authors or sources. Clinical trial results published after January 1, 2013, were also included in my search. Seminal publications, regulatory guidelines and authoritative websites were included in the literature review. Additional

publications were identified, by reviewing the citations within identified articles. Abstracts, methods and possible relevant target populations were also reviewed.

Theoretical Foundation

The quantitative theoretical framework is a set of constructs that indicate the relationship between the variables included in the study hypothesis or hypotheses (Creswell & Creswell, 2014). The variable relationship and how they interact with each other can be described using a theoretical framework, where the theory will provide a possible explanation of the relation of the variables of interest. The system could be considered an organism where there are a finite number of interacting variables rather than isolated parts. The general systems theory developed by Kenneth Boulding in 1956 is the study of systems and how the interrelated and interdependent parts interact with each other (Boulding, 1956). The underlying assumption of the theory is that as one component of the system changes, this change will affect other components of the system.

Systems theory has been used in a variety of fields including social work (Forder; Warren, 1998), career development (Patton & McMahon, 1999, 2006), business management (Chikere & Nwoka, 2015; Mele, Pels, & Polese, 2010), language learning (De Bot, Lowie, & Verspoor, 2007; X. Huang, Acero, Hon, & Reddy, 2001) family systems therapy (Becvar & Becvar, 2017; Knudson-Martin, 1994) and community development (Lerner, Almerigi, Theokas, & Lerner, 2005). In each field of research, the theory provides structure as to how the components impact each other and when one aspect changes, the change can impact the entire system. For example, in social work,

systems theory can help explain an individual's behavior based on a multitude of interrelated components. An individual's family life, community, social structure and the individual themselves all impact how the individual may respond to a situation. In the research that uses systems theory, there is a system of components that are interrelated and dependent on each other. An alteration of one component can positively or negatively change the other components of the system.

In this study, I assessed the association between early phase design methods, specifically traditional versus adaptive and how the design method was associated with the late stage study outcome (positive, equivalent or negative). The underlying hypothesis is that when adaptive methods are used, increased and high-quality information is gained, leading to improved results in late stage studies. Adaptive methods may not be as effective in every type of cancer, treatment classification, or experimental drug classification, but the impact of the study characteristics were assessed when adaptive methods are used versus not.

Within systems theory, changing or altering one component can impact another (Boulding, 1956). This characteristic is true in clinical trial development. For example, changing one classification, such as target population, will likely impact the type of treatment as well as drug classification, as cancer is a diverse number of diseases with a variety of treatment paths due to a variety of cell proliferation paths (Ardies, 2014; Lonial & Nooka, 2016; National Cancer Institute, 2015).

Within systems theory there is a feedback loop, which is a process in which the system uses information generated within the system (Boulding, 1956). The process of

clinical research development uses early phase study results to provide input on future study development. Further, late stage study results can also impact future early phase studies, as the experimental treatment may be assessed in other indications or cancer subgroups. However, the time between the studies also needs to be considered. For example, studies conducted 20 years ago may have minimal impact on current studies due to changes in standard of care as well as differences in the target population over time. In contrast, an ongoing study may not be impacted by a study that has just reached completion due to the proximity in time. However, the ongoing study may be dramatically impacted by the recently released results due to shared study characteristics, such as similar treatment or targeted population. If the results of the recent study are negative and the shared study characteristic is target population and treatment, the ongoing study may be stopped for ethical reasons. This aspect of systems theory is considered to be self-correcting as the components of the system react from information provided by other components of the system(Boulding, 1956).

The hypothesis of this study was that the component design method (adaptive or traditional) may be associated with the downstream results of the later stage study. The hypothesis was based on the idea that by using the innovative design methods, the late stage outcomes may be improved due to the quality and increased quantity of the information gained. Figure 1 provides an outline of systems theory and how the theory relates to clinical research and the assessment of adaptive and traditional methods.

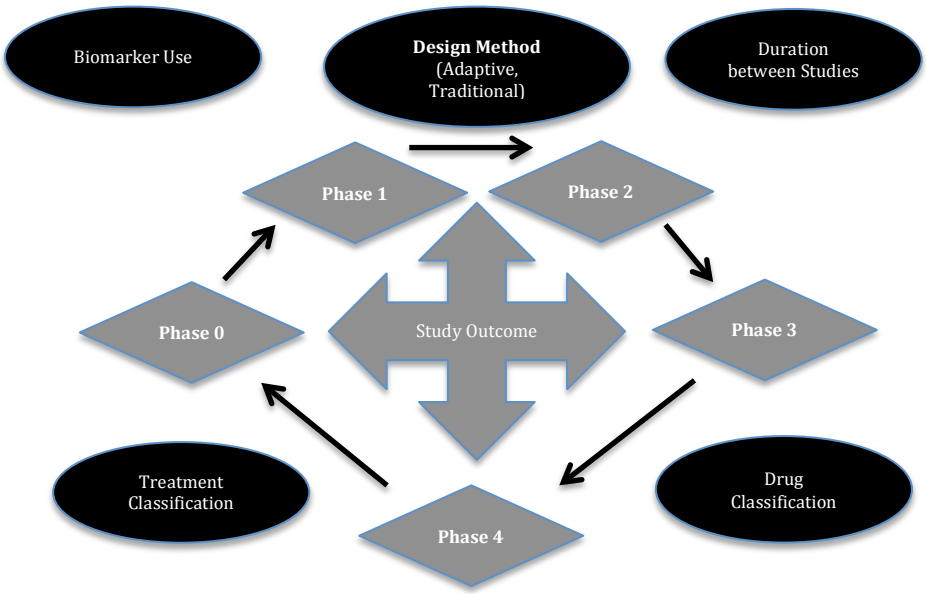

Figure 1. Systems theory applied to oncology clinical trial research and the association of early stage design methods to late stage outcome results.

Conceptual Framework

Researchers have noted that while randomized clinical trials (RCTs) are the gold standard of treatment development, outcomes of complex interventions need to be systematically, interactively developed and assessed (Campbell et al., 2000). The Framework for Design and Evaluation of Complex Interventions (DECI) provides guidance on the integrated evaluation of complex interventions (Campbell et al., 2000). The framework was developed to assess complex research interventions with various interconnected components. The information gained within a study as well as the study results are integrated within and across every phase in an iterative fashion (Campbell et al., 2000). Due to the complexity of treatment and interventions in clinical trials, learning across and within studies is critical to continue learning and improving clinical trial

outcome as well as assessing unexplored endpoints (Geifman & Butte, 2016). Further learning in and across studies with positive and negative findings will aid in advancing the field (Butterfield et al., 2010).

The variables included in the model indicate the complexities related to the diverse interventions, specifically the treatment of cancer in clinical trials. Important factors to consider in oncology clinical research include target population, treatment and drug classification, study sponsor as well as the duration between early and late stage study. The phases of my study (early phase: 0, I and II) are indicated in Figure 1 and Figure 2. The late stage studies (phase III and IV) are also included in Figure 2. The variables in the model and the quality of the results are being used to improve the study development of quality late stage studies. In each of the research questions and associated hypotheses for this study, I assessed whether the quality of the results and information was improved if adaptive methods are used.

While the authors of the conceptual framework indicate that a single intervention such as a drug is not complex (Campbell et al., 2000)., I assessed multiple drugs for multiple indications, which is indeed complex One factor that makes oncology research complex is in multicountry studies, standard of care (SOC) and best supportive care may not be standard across locations thus integrated analyses should be carefully conducted (Organisation for Economic Co-operation and Development [OECD], 2013). In addition, standard of care is likely to change over time or during a study; thus, differences with respect to historical controls in randomized studies may be attributed to the SOC rather than the treatment of interest, leading to biased estimates or improvements not

attributable to the experiment treatment (Viele et al., 2014). While these variables were not included in my analysis, researchers should consider these aspects when integrating results of clinical research.

Complex interventions that utilize the DECI framework are provided by the authors, such as service delivery and organization for stroke units as well as community based programs to prevent heart disease (Campbell et al., 2000). Researchers have noted the need for well-designed and implemented studies, as well as the evaluation of the studies through meta-analysis and long-term programs, utilizing the DECI methods of systematic review and iterative information gain (Grol, 2001). Researchers using adaptive methods may lead to better designed studies leading to improved information that can be used for integration into other studies. In addition, systematic reviews of clinical trials and health related data are critical for improving methods as well as treatment for the targeted populations (Griffiths et al., 2006; Grol, 2001; Kroeze et al., 2006; Steinert et al., 2006). This research included oncology studies randomly identified in the ClinicalTrials.gov database. A stratified random identification of studies was conducted to allow for increased power for sub-population comparisons. The study was developed to assess the association of quantitative methods related to study design and the results of late stage studies.

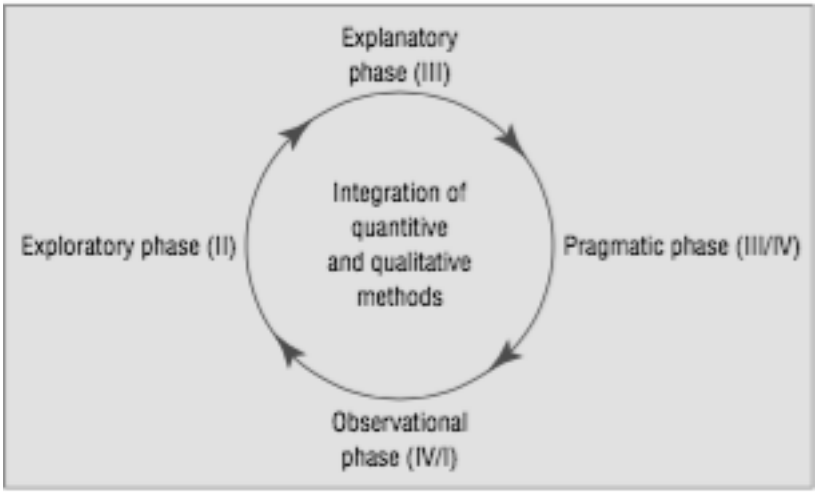

Figure 2. The iterative process of the Framework for Design and Evaluation of Complex Interventions (DECI). Reprinted with permission from "Framework for design and evaluation of complex interventions to improve health" by M. Campbell et al., BMJ: *British Medical Journal, 321*(7262), 694-696.

Due to the sequential nature of oncology clinical research, an iterative approach should be consistently applied and interventions re-examined as needed information is collected (Campbell et al., 2000; Geifman & Butte, 2016). Thus, the early phase studies are not just providing guidance for late stage studies, but are also providing input for other ongoing and future early phase studies, that use the same treatment or is directed toward the cancer indication. The development of clinical research is iterative to ensure that new information is captured and utilized for the development of early and late stage studies. My research does not include the qualitative methods that are mentioned in DECI, however, the qualitative component of the research could be conducted in the future and is included in the future research section of my dissertation. Consideration of a survey to assess the resistance to the adoption of adaptive methods and sent to oncology

researchers could be conducted in the future to assess qualitative aspects related to adaptive methods.

The focus of this research was to quantitatively assess adaptive methods versus traditional methods used in oncology clinical trials. My dissertation used early study research, which is then used to develop late stage studies. The interrelationship of the phase of treatment is a key component of DECI.

Literature Review

Cancer – Epidemiology and Heterogeneity

Cancer is defined to be the uncontrolled division of cells that may lead to the invasion of nearby or distal tissue (National Cancer Institute, n.d.). Cancer is the second leading cause of death for men and women in the United States (Centers for Disease Control and Prevention, 2017), with the most commonly diagnosed cancers in 2017 projected to be breast cancer, lung and bronchus cancer, prostate cancer and colon cancer (National Cancer Institute, 2017). More than 1.6 million cases of cancer were estimated to be diagnosed in the United States and over a half million individuals were reported to have died from the disease in 2016 (National Cancer Institute, 2017). While deaths from cancer have decreased, diagnoses of some cancers have stabilized and the incidence of most cancers continue to increase (National Cancer Institute, 2017). Further, individuals that are living beyond the cancer diagnosis are expected to continue to rise (National Cancer Institute, 2017), thus the need for treatment development that reduce the long term impact on quality of life. Even with reductions in smoking rates, cancer rates continue to increase in the United States as well as across the globe as a result of a variety

of factors including an aging population (Centers for Disease Control and Prevention, 2017). With the increasing demands on healthcare and treatment related to cancer in the United States, effective and safe treatments need to be developed and identified through clinical trials with therapeutic dose levels and targeted populations.

Cancer can originate in the bone, internal organs, central nervous system, blood or bone marrow and can spread through the body via the blood or lymphatic system (National Cancer Institute, n.d.). Due to the variety of origination of disease as well as the stage of disease at diagnosis and the risk factors associated with the individual diagnosed such as age, environment, or genetic biomarker, cancer is not one disease; but a constellation of diseases (Ardies, 2014; Lonial & Nooka, 2016). As a result of different cell proliferation mechanisms related to cancer, drug treatment has to be developed to address the cancer disease diversity. Due to the diversity of the diseases as well as treatment development, target populations also need to be considered and identified in protocol development.

Though cancer has been reported to be a single disease, cancer is actually a complex set of diseases (Butterfield et al., 2010). Often the reporting of the cure for cancer is reported in the news. However, researchers know that cancer is not a single disease, but a classification of hundreds of diseases (American Association for Cancer Research [AACR], n.d.). Cancer is a multitude of diseases with different mechanisms leading to cell proliferation (National Cancer Institute, 2015). As a result, treatment for each cancer likely needs a different plan of action to address the complexity of the diseases (National Cancer Institute, 2015).

The same cancer in children, young adults versus elderly may have different characteristics leading to different modes of treatment (Boissel et al., 2003; DeAngelo et al., 2015). Within a single type of cancer, there are a variety of disease characteristics that lead to variation of treatment effectiveness and response (Carey, Winer, Viale, Cameron, & Gianni, 2010; Kim, Ueda, Naka, & Enomoto, 2012; Lonial & Nooka, 2016). Due to the disease variation, to date, there is no single treatment that has been found to be effective on all cancers. A treatment found to be effective in breast cancer, may not be effective in pancreatic cancer. A treatment found to be effective in HER2+ breast cancer, may be ineffective in HER2- breast cancer. Due to the complexity of a single type of cancer as well as the diversity of treatment response across cancers, a variety of treatment options need to be developed and assessed (Barker et al., 2009). Cancer treatments need to continue to be developed in an effective manner to determine efficacy and safety, as well as identify the appropriate target population based on the treatment mechanisms and patient characteristics. Adaptive studies can aid in identifying target populations early in the process (Scher, Nasso, Rubin, & Simon, 2011). The complexity of cancer needs to be addressed in study design.

Cancer Risk Factors

As the disease itself varies, there is some consistency in risk factors related to cancer such as diet and obesity (Burger et al., 2013; Kampman, Vrieling, van Duijnhoven, & Winkels, 2012; Ross, 2010; Vera-Ramirez et al., 2013), tobacco use (Burger et al., 2013; Gillison et al., 2012; Maxwell et al., 2010; Warren & Cummings, 2013) and exposure to other chemicals such as asbestos or particle pollutions (Berman &

Crump, 2008; O'Reilly, Mclaughlin, Beckett, & Sime, 2007). Occupational exposures to chemicals and particulates have been found to be related to increasing cancer rates (Alavanja, Hoppin, & Kamel, 2004; Burger et al., 2013; Purdue, Hoppin, Blair, Dosemeci, & Alavanja, 2007). While an individual's behavior can impact their risk of cancer, increasing age (A. Y. Chen, Jemal, & Ward, 2009; Jemal, Siegel, Xu, & Ward, 2010), as well as hormones and genetics (Burger et al., 2013; Vera-Ramirez et al., 2013) have been shown to increase the risk of certain cancers. Biomarkers can also indicate the aggressiveness of the cancer (Gravdal, Halvorsen, Haukaas, & Akslen, 2007; Lim et al., 2009), as well as the appropriate treatment path that the medical staff and patient should take for their disease (Barker et al., 2009).

Researchers continue to identify mechanisms of cancers in order to develop treatments to impact the disease path (Barker et al., 2009; Gravdal et al., 2007). Further, researchers also address risk factors in clinical trials through stratification factors, in an attempt to balance the population across the randomized treatments as well as classify patient baseline characteristics (Barker et al., 2009). This classification allows for better understanding of the patient population as well as the assessment of the impact of the treatments on a specific population. Once again, the complexity of cancer is addressed through the identification of risk factors and the appropriate treatment is prescribed.

Treatment Development and Funding for Cancer

All cancer clinical trial development, progresses through a similar process called phases of treatment development, regardless of the cancer type. After the pre-clinical assessments including animal models, there are, in general, three phases of studies that

are conducted on human beings; phase I, phase II and phase III. Each phase of development has specific and detailed objectives as well as sample size and provides information for later phase studies or studies in other target populations.

Phase I studies are conducted on humans where the objective is to assess safety as well as determine the maximum tolerated dose (National Institute of Health, 2016). Unlike other nononcologic indications, cancer patients rather than healthy volunteers are often used in phase I cancer clinical trials due to limited treatment options for those with the disease (Salzberg, 2012). Phase I cancer trial sample sizes have increased over the years, with the median number of patients (Q1, Q3) estimated to be approximately 55 patients (36, 80) (Dahlberg, Shapiro, Clark, & Johnson, 2014). The objective of cancer phase I studies is to assess safety and pharmacokinetics (the study of the movement of the drug in the body) and pharmacodynamics (the study of the effects and mechanisms of drug action) related to the specified dose can also be assessed (National Institute of Health, 2016). Efficacy is also usually estimated with a short time framed endpoint, such as treatment response.

Phase II studies are usually larger than Phase I studies with a continued focus on safety, however, efficacy is more formally assessed. (National Institute of Health, 2016). There may be multiple efficacy endpoints, of which, the primary is likely a short-termed endpoint once again. While the sample size can vary, sample sizes of less than 100 patients are most common (J.-H. Huang et al., 2015). These studies also may be randomized, but not necessarily comparative.

Phase III studies tend to be larger in sample size allowing for the comparison of multiple treatments (National Institute of Health, 2016). Phase III studies are often randomized studies to allow for the comparison of the new treatment versus the standard treatment of which the sample size can vary from approximately 100 patients to thousands (Salzberg, 2012). Phase III studies tend to focus on efficacy and likely includes other secondary objectives including safety and patient reported outcomes while controlling operational characteristics such as Type I and Type II errors (Ioannidis, Hozo, & Djulbegovic, 2013; National Institute of Health, 2016). Combined studies, otherwise known as seamless designs such as phase I/II or phase II/III can also be developed with the objective of attaining the study answers and treatment development more quickly (National Institute of Health, 2016).

Phase 0 and Phase IV are less common studies that can be incorporated into treatment development. Phase 0 studies are considered exploratory and allow for the exploration of agents to be assessed in phase I studies (Doroshow & Parchment, 2008). Phase IV studies are also less common, but are conducted to assess long term safety and efficacy after treatment regulatory approval (National Institute of Health, 2016). Information gained at each phase of treatment provides guidance for the related studies conducted subsequently. Multiple phases can also be combined in a single study.

The report developed through BioMedTracker tracks the clinical development and regulatory history of investigational drugs. The researchers found the probability of success going from phase I to phase II is 63.2% and from phase II to phase III is 30.7% for all modalities (Biotechnology Innovation Organization [BIO], 2016). While the

researchers assessed fourteen different modalities, oncology had the lowest likelihood of regulatory approval starting from phase I studies with 5.1%, even though almost 31% of the drug development program transitions were in oncology. Innovative methods need to be adopted in oncology to combat low clinical research success.

The cost and time consumption of oncology drug development has increased over the years, without a corresponding treatment success rate (Berry, 2012; S.-C. Chow & Chang, 2008; Christopher S. Coffey, 2017). Researchers assessed the cost of ten recently approved cancer drugs and the median cost for development of the single drug was approximately $648 million (Prasad & Mailankody, 2017). The time from development to approval is also lengthy with medians ranging from 58.8 to 93.5 months dependent on submission types (Jardim, Schwaederle, Hong, & Kurzrock, 2016). With the multi-phase process, time and finance invested in clinical trials as well as the impact to current and future patients, improved methods including study design may lead to a higher success rates.

Adaptive and Traditional Design Characteristics

Ambiguity of adaptive design definition. Each phase of oncology studies has a specific objective or objectives likely related to efficacy or safety. In phase I studies, the focus tends to be safety, specifically determining the maximum tolerated dose (MTD) that will be used for the recommended phase two dose (RP2D). While confirmatory adaptive designs have been researched for over 25 years (Bauer, Bretz, Dragalin, König, & Wassmer, 2016), the methods became more widely known and utilized five years later (Bauer et al., 2016). During this time period however, the definition of adaptive designs

has been less clear. The ambiguity of the definition of adaptive designs has been a significant barrier with respect to the method adoption (Christopher S Coffey et al., 2012; Dimairo et al., 2015).

Definition of adaptive design. Adaptive designs, for the purpose of this analysis is clinical study designs that use accumulated information (data) of the study, to modify aspects of the study as the study continues (Christopher S Coffey et al., 2012). Examples of adjustment include underlying assumptions include the mean, variance or the sample size. The most common adaptations in oncology include stopping for futility or safety, adjustments in dosing or treatment assignments, identifying a target population with an effective treatment response or seamless multi-phase studies (Berry, 2012; Reitsma, 2015). In adaptive designs, the adjustments to the study are pre-defined, thus are not made on an ad hoc basis. The changes are not the result of inadequate planning, but are preplanned components of the design (Kairalla et al., 2012).The adaptations can be utilized on a single endpoint or a combination of endpoints that have been identified as predictive in nature (Berry, 2012). Studies can also be adapted on nonprimary endpoints such as survival post progression (SPP), progression free survival (PFS) for a study with overall survival (OS) as the primary endpoint (Berry, 2012). The adaptive design definition used for this study is consistent with the Pharmaceutical Research and Manufacturers of America (PhRMA) Working Group (Gallo et al., 2006)

Traditional design characteristics. For this study, traditional study designs will be prespecified rule-based designs such as the 3+3 design. The 3+3 design methods were introduced in the 1940s as a method to systematically escalate the treatment dose while

monitoring the safety of the treatment (Hansen et al., 2014). In general, the initial set up patients (usually three) are enrolled at a subclinical treatment dose and are assessed for dose limiting toxicities (DLT) for the first cycle of treatment. If less than a prespecified number of patients experience a DLT, the next set of patients, usually three patients, are enrolled at the next highest dose. If once again, a prespecified number of patients experience a dose limiting toxicity, then an additional set up patients, usually three, are enrolled at the current dose. Dose escalation can only occur if there are less DLTs than expected. The dose escalation is often predefined and frequently based on Fibonacci series so that dose increments are smaller at higher doses.

Limitation of traditional design. However, the maximum tolerated dose (MTD) may not be attained due to a lack of adverse events defined as DLTs being reported. For treatments that are cytostatic rather than cytotoxic, the toxicity rate may increase based on drug accumulation rather than the incremental increase of the dose level. Researchers have noted that the MTD is only identified in approximately 30% of trials (Ananthakrishnan et al., 2017; Hansen et al., 2014). By design, there are excessive dose escalations and pauses between steps, leading to longer study duration (Le Tourneau, 2009). Due to the escalation design and the conservative methods, a high proportion of patients may be treated at suboptimal effective dose levels. Unfortunately, the 3+3 methods have not been statistically supported and have been shown to be ineffective through simulations and oncology clinical results (Christopher S Coffey et al., 2012; U. S. Food and Drug Administration, 2011; Hansen et al., 2014). As such, the recommended

treatment dose may be erroneous impacting the treatment of future patients. Improved study design methods need to be developed and adopted.

Traditional versus adaptive design assumptions and operating characteristics. When using the 3+3 design, there are underlying assumptions based on the dose and the response. Specifically, the 3+3 design assumes that as the dose increases, the related toxicities also increase. Researchers have noted that even in drugs where the assumed dose-toxicity relationship does not hold, the design methods are still utilized 60% of the time (Hansen et al., 2014). These methods are often used in early stage oncology studies and perhaps may be the result of the poor treatment approval with a failure rate as high as 66% in results reported from 2003 to 2010 (Berry, 2011). Researchers recommend including simulations of dose-toxicity operational characteristics when fixed designs are being used. This may improve understanding of assumptions related to the traditional design method use (Bornkamp et al., 2007).

Using traditional methods, the success of the study is dependent on the original underlying assumptions. Unlike traditional methods, adaptive designs provide a path to address uncertainty during the original design phase. Adaptive design methods use information already accumulated on the trial allowing flexibility, which may increase the success of the study. In traditional designs, underlying assumptions are often based on historical studies which are used to determine the sample size (C. H. Brown et al., 2009). However, when using adaptive methods, the initial design may utilize the same assumptions as the traditional design, but during the study, the underlying assumptions can be assessed in a blinded or unblinded fashion and adjustments to the design can be

made. For example, when using the adaptive methods such as sample size re-estimation, the underlying assumptions such as endpoint variability can be assessed during an interim review and the sample size adjusted accordingly (Chuang-Stein anderson, Gallo, & Collins, 2006).

Reluctance to adopt adaptive methods. With the introduction of the adaptive methods and the indication that they are more effective than traditional methods (Berry, 2011; Christopher S. Coffey, 2017), researchers unfortunately continue to be reluctant to adopt the methods. Reasons for reluctance are diverse but include a lack of understanding of the methods (Kairalla et al., 2012), lack of education and access to case studies (Dimairo et al., 2015), increased time for planning (Collinson et al., 2012; Dimairo et al., 2015) and the need for additional software and infrastructure (Kairalla et al., 2012). Researchers also note the limitations of access to off the shelf packages identified as another barrier of adoption (Bornkamp et al., 2007). When the adaptive methods are used, they are most commonly used in phase II studies (Hatfield et al., 2016). Expansion of their use should be considered for phase I studies.

Researchers have found that studies initiated from 1991 to 2006, 98% were designed using traditional methods (Rogatko et al., 2007). In studies initiated between the years 2000 and 2014, while adaptive methods have been increasing in use, a dramatic uptake in use of methods was not observed (Hatfield et al., 2016). During this time period, in all clinical trials, only 143 of 573 (25.0% of total studies identified) nonphase I adaptive studies were included in their analysis (Hatfield et al., 2016). Barriers related to

the adoption of adaptive methods need to continue to be identified and addressed for future as well as current researchers.

Advantages and disadvantages of traditional methods. While traditional methods have limitations, there are some advantages to the methods as well. Researchers have noted that the traditional design methods are simple to use and understand leading to their continued use (Ananthakrishnan et al., 2017; Buoen, Bjerrum, & Thomsen, 2005; Le Tourneau, 2009). While the methods may not have clinical success or statistical justification (Bornkamp et al., 2007; Buoen et al., 2005), the methods are commonly used due to preference, habit and conservative dosing (Buoen et al., 2005). Based on the dismal success rate of studies in oncology, however, a shift in the clinical research development process needs to occur.

The failure rate of oncology studies, submitted between 2003 to 2010 was 66% (Berry, 2011), indicating that innovative methods need to be developed and adopted in clinical trials (Berry, 2012; S.-C. Chow & Chang, 2008; Christopher S. Coffey, 2017). The traditional methods also fail to identify the maximum tolerated dose in approximately 30% of trials (Ananthakrishnan et al., 2017), which may be due to the methods being underpowered (Butterfield et al., 2010). In addition, the dose response curves are usually not estimable due to the low sample size and the study being underpowered once again (Kairalla et al., 2012). The traditional methods also have a conservative dosing trend, thus excessive patients tend to receive sub-therapeutic doses without the opportunity for intra-patient dose escalation (Ananthakrishnan et al., 2017; Butterfield et al., 2010; Hansen et al., 2014; Le Tourneau, 2009). Historical controls for

parallel designs also have limitation due to population and standard of care changes over time that cannot be controlled within study parameters (C. H. Brown et al., 2009). While traditional methods are simple to understand, their inaccuracies and limitations are a detriment to current and future patients that need to be addressed.

Often in early stage traditional studies, clinical decisions to continue the treatment development are based on toxicity alone (phase I) or rapid response efficacy data (phase II) (Butterfield et al., 2010). Decisions to continue or stop the study are based on the current dose only and not the accumulative information of all treatment doses assessed in the study (C. H. Brown et al., 2009). The accumulation of information in the adaptive trial, guide the flexibility of the design, which may contribute to the success of the studies. The additional information gained is leveraged for improved study operating characteristics (Kairalla et al., 2012; Reitsma, 2015). Unlike traditional designs, adaptive studies can use historical data for the initial design, but also use the current study information for adjustments to the underlying assumptions.

Early phase studies assess the treatment for safety as well as identify the maximum tolerated dose. Efficacy is more thoroughly assessed in later stage studies. However, due to the early phase traditional designs being underpowered (Butterfield et al., 2010), the accuracy of these studies can lead to negative downstream impact on subsequent trials. Further, when the differences in the DLT rates are smaller, the accuracy of the MTD selection is reduced when traditional methods are utilized (Ananthakrishnan et al., 2017). Often when adaptive methods are utilized, simulations are conducted so the entire clinical team can understand the operating characteristics under a variety of

assumptions. However, when traditional methods are used, trial performance on identifying the correct dose are often not assessed (Bornkamp et al., 2007) but should be.

Traditional designs use underlying assumptions from previous studies in order to develop the current study. While this may be the case for both traditional and adaptive designs, adaptive designs can modify prespecified characteristic(s) in order to reflect the cumulative data. While traditional studies depend on theoretical underlying behavior often assumed from other studies, adaptive studies rely heavily on simulations to increase knowledge of trial behavior, study characteristics and possible risks (Reitsma, 2015; Viele & McGlothlin, 2017). More simulations should be conducted when using traditional methods so that there is a greater understanding of study characteristics and possible outcomes as well (Reitsma, 2015; Viele & McGlothlin, 2017). If operating characteristics were assessed when traditional methods were being utilized, perhaps increased knowledge with respect to their limitations would be more understood.

Additional knowledge with respect to the underlying assumptions of traditional methods, specifically the 3+3 also need to be addressed. The intended use of the 3+3 assumes that toxicities increase with the dose. However, for biologics this assumption may not be valid (Le Tourneau, 2009). This trend was also found even in drug classifications of molecularly targeted agents (MTA) of which more novel design methods should be utilized due to the assumption related to dose and adverse events. In MTAs, toxicities tend to develop as the result of accumulation of drug rather than the actual dose. However, 60% of those studies utilized conventional 3+3 design methods

(Hansen et al., 2014). Once again, traditional operating characteristics as well as underlying assumptions should be assessed before the methods are adopted.

Need for innovation. Learning from other studies is critical, but also using the maximum information from within a study is also important. Using updated design methods, such as adaptive designs, which are designed to use accumulative information within the study, rather than the single dose level data used in traditional designs. The adaptive methods have been shown to be more effective at identifying the appropriate treatment (Berry, 2011; Christopher S. Coffey, 2017; Christopher S Coffey et al., 2012; Dimairo et al., 2015; Hatfield et al., 2016), as well as more efficient with respect to precision, power and controlling the type I error level (Bornkamp et al., 2007). Researchers have also noted that adaptive design methods can more quickly identify ineffective treatments reducing the risk to the targeted patient population (Christopher S Coffey et al., 2012).

When using traditional methods, one underlying assumption is that as the dose increases, so does the rate of toxicities. However, in newer noncytotoxic drugs such as molecular targeted agents and cytostatic drugs, the dose toxicity relation is not maintained. As such, researchers need to understand the characteristics of their study drug as well as the adopted design methods (Kairalla et al., 2012). Researchers also need to understand the benefit of using cumulative information within a study rather than only information from a specific dose. As a result of using the accumulated data, adaptive trials tend to eliminate failures earlier (Reitsma, 2015), thus patients are less likely to be exposed to sub-therapeutic doses. Once again, the positive impact of adaptive methods

can be seen in the early phase I-SPY2 (Investigation of Serial Studies to Predict Your Therapeutic Response with Imaging and moLecular Analysis 2) early phase program. The I-SPY2 studies that successfully proceed out of the phase I study, are expected to have an 85% success rates in confirmatory studies (Reitsma, 2015). The Critical Path Initiative was established to improve and facilitate innovative discussions (U. S. Food and Drug Administration, 2004). Biostatistics and associated methods are one of the initiatives with the objective of improving tools, methods and study designs for drug development. The studies, I-SPY1 and I-SPY2 were the result of these initiatives (Parekh et al., 2015). Innovative methods knowledge needs to continue to expand so that the methods are utilized appropriately and more frequently in clinical research.

While learning from within a study is critical (S.-C. Chow & Chang, 2008; Kairalla et al., 2012), learning from outside the study is also important. Researchers have suggested that access to analysis code and output will help with the design of future studies (Dimairo et al., 2015). The I-SPY2 adaptive study is an excellent example of shared data and results from study to study as well as shared information within the study (Barker et al., 2009). Shared information of positive and negative results should occur more frequently to help improve oncology research results and reducing the chances of repeating failed research. The ClinicalTrials.gov database is a good start to assess shared design and initial results, however, improved data structure is necessary that reduces free text fields and collects additional response data as researchers have suggested (Hatfield et al., 2016). Unfortunately, the registry databases have not completely eliminated publication bias related to study results (Dwan, Gamble, Williamson, & Kirkham, 2013).

Sharing of information across studies, regardless of outcome will be a critical step forward in improving the results of clinical trials thus increasing effective treatment options for patients.

Both traditional and adaptive methods can be utilized in adaptive clinical trials using strengths of each method. However, researchers need to understand the strengths and operational characteristics of each method so they can be appropriately utilized (Berry, 2011). Researchers have noted that while adaptive designs may not always be recommended (Korn & Freidlin, 2017), when studies include more than two treatment arms, adaptive methods do appear to be more efficient and effective (Berry, 2011). Of the drugs that were approved by the FDA from 1992-2008, 21 of 25 (84%) used the 3+3, where more than 50% had 6 or more dose levels (Le Tourneau, 2009). Early phase oncology studies are ideal for adaptive design methods as the determination of the maximum tolerated dose is critical, thus the emphasis on controlling type II errors (false positive) need to be considered (Kairalla et al., 2012). Type II errors can result in future patients receiving too low a dose, which likely will impact efficacy in the target population. When there are a high number of dose levels in determining the MTD within a study, researchers once again need to assess innovative design methods and determine if the methods are more appropriate for their study.

Challenges for adaptive design adoption and overcoming barriers. Before adaptive design methods are adopted, a clear definition of adaptive methods needs to be established and understood. This has been a significant barrier related to the acceptance and adoption of adaptive designs (Christopher S Coffey & Kairalla, 2008; Christopher S

Coffey et al., 2012; Dimairo et al., 2015). Researchers have noted the ambiguity of what adaptive design really means and has been and continues to be a barrier for adaptive design method adoption for more than ten years. Even reviewing the ClinicalTrials.gov database, a traditional dose escalation design used the term ‘adaptive’ to describe the study design, though based on publication of the results, the study appears to be a traditional design (NCT02281786) (Panza et al., 2016). Further, unplanned changes to the design is not considered adaptive designs (Dimairo et al., 2015). Poor planning, leading to the need in the design change is not an adaptive design.

Regulatory considerations related to innovative designs such as adaptive methods, need to be addressed before design methods are utilized. Researchers need to increase their knowledge in understanding the innovative methods as well as the regulatory requirements associated with adaptive methods (S. C. Chow et al., 2012; Christopher S Coffey & Kairalla, 2008; Dimairo et al., 2015). Knowledge associated with regulatory requirements appear to be growing, thus regulatory concerns are becoming a reduced limitation (Dimairo et al., 2015). Recommendations from regulatory agencies include satisfactory simulations reflecting multiple situations (Viele, 2017), in order to assess the operating characteristics of the design and the behavior under a variety of situations (Miller et al., 2017). Researchers have also published case studies, best practices and adaptive design characteristics to increase knowledge as well as increase and improve adaptive design use, particularly for less well-understood adaptive designs (Table 2) (He et al., 2017; Miller et al., 2017). Regulatory bodies as well as researchers have voiced their concerns on controlling alpha spending, confidence interval and *p*-value estimates

due to the study adaptations (S.-C. Chow & Chang, 2008), thus study operating characteristics should be reviewed by the clinical study team to increase understanding as well as provide documentation to regulatory bodies. The challenge recommendations for adaptive designs are similar to study conduct using traditional designs, though statistical methods may have been more recently developed. Like traditional methods, appropriate statistical analyses methods need to be utilized to reflect the study design to reduce the introduction of bias as well as control for type I and type II errors (He et al., 2016).

Traditional methods have been used since at least the 1940s (Hansen et al., 2014) and are relatively simple to develop and utilize. There has however been an increase in use of adaptive methods over time (Bauer & Einfalt, 2006; Hatfield et al., 2016). To use adaptive methods, an increased amount of time for planning is needed and must be considered in the timelines prior to initiation of development (Dimairo et al., 2015; Kairalla et al., 2012). Funding of this additional time also needs to be considered and addressed prior to study development (Kairalla et al., 2012). Aside from additional initial funds, infrastructure including programming and randomization must be in place during protocol development (Christopher S. Coffey, 2017; Christopher S Coffey & Kairalla, 2008; Hansen et al., 2014; Hatfield et al., 2016). While the FDA has allocated funds to facilitate innovative discussions and method utilization, in order to access the funds, the study must be designed before you can apply for the grant (Christopher S. Coffey, 2017).

Table 2

Best Practices and Challenges for Less Well-Understood Adaptive Designs

Challenge	Best practice
Type I error control	Appropriate statistical techniques related to planned analysis as suggested by Wassmer & Dragalin (Wassmer & Dragalin, 2015), Chen, DeMets & Lan (Y. Chen, DeMets, & Gordon Lan, 2004) and Schmidli et al.(Schmidli, Bretz, Racine, & Maurer, 2006).
Data Monitoring Committee (DMC) review process and the role of the DMC	DMC members should have the expertise, experience to review according to the adaption plan. Also should restrict interim results knowledge to a small decision-making group.
Statistical bias related to treatment effects estimates	Appropriate simulations to understand the potential for bias in specific situations.
Subject heterogeneity across study stages	Minimize protocol amendments. Enroll across regions or sites in similar timeframes.
Potential for making decisions based on highly variable and unreliable interim results	Appropriate timing of interim analysis (follow up and reducing variability). Targeting interim analysis to occur with 50 to 75% information or sample size.
Potential for overrun of subjects being recruited	Planning prior to analysis is critical including continual data cleaning, timing of interim analysis, programs and firewall ready.
Issues with seamless phase II/III trials	Consideration between accrual speed and endpoint assessment.

Adapted from "Addressing challenges and opportunities of "less well-understood" adaptive designs. " by He, W., Gallo, P., Miller, E., Jemiai, Y., Maca, J., Koury, K., ... & Lin, M., 2017, *Therapeutic Innovation & Regulatory Science*, *51*(1), 60-68.

Aside from adaptive design methods being adopted and understood, other clinical research areas of expertise are needed such as data management. Data management processes need to be addressed to reduce bias related to logistical challenges (Dimairo et al., 2015). For example, the timely reporting of data including dose exposure, treatment response and related adverse events are critical for the quality assessment of a dose level. Further, research institutions may need to have a dedicated team for the development of

adaptive studies, as researchers have indicated that an adaptive working group does appear to help with respect to expertise and knowledge dissemination (Morgan et al., 2014). Funding gaps also need to be addressed, as there has been a lag in adaptive design adoption in publicly funded research (Dimairo et al., 2015; Hatfield et al., 2016).

While there are barriers that need to be overcome to address the adoption of adaptive design methods, there are some simple solutions that researchers have suggested. Adaptive methods education in the universities is necessary but not just exclusively for statisticians (Dimairo et al., 2015). Education opportunities including hands-on experience, however, need to expand beyond the university to address the needs of current researchers. In addition, the justification and explanation of the trial options to the researchers to have a better understanding of traditional versus adaptive methods is critical (Viele & McGlothlin, 2017). The adoption gap between public and private funded research confirms the need for sharing of resources through cross funding working groups such as the Drug Information Association's (DIA) Adaptive Design Scientific Working Group (ADSWG). Further, sharing of initial design related outputs could also increase expertise across the field (Morgan et al., 2014).

Limitations of clinical research. Oncology clinical trials are developed in general sequential order to utilize information from the previous study. Using data from studies cumulatively is becoming more common through the use of meta-analyses, these methods should be utilized and adopted consistently across clinical research, phase and indication. Unfortunately, in clinical research, when a study fails, often results are not published (Hopewell, Loudon, Clarke, Oxman, & Dickersin, 2009). Only the researchers

involved in the study itself may fully understand and learn the reasons for the failed study due to the reduced likelihood of the results being published (Chapman et al., 2017; Gluud, 2006; Raghav et al., 2015). This knowledge can be applied to other internal studies, however, the information will not likely be publically shared with external researchers. Researchers have noted the phenomenon of publication bias where positive studies are more likely to be accepted for publication in comparison to negative studies (Gluud, 2006). While positive studies are critical for patient treatment, assessing the results and study design of negative studies is also critical in improving the treatment related to oncology. Researchers have noted that unplanned endpoints with positive results are more frequently reported than negative results in abstracts (Raghav et al., 2015). Researchers have categorized negative trials and research no longer of interest in the same category for reasons not published (Chapman et al., 2017), which is a concern. Researcher assessment and information exchange of positive and negative studies will aid in advancing the field.

Researchers have noted that appropriate endpoints need to be used and developed in clinical research (Collinson et al., 2012; Fleming, 1996). The appropriate endpoint also needs to be selected based on all study specific information such as treatment mechanism, type of cancer and planned follow up (Collinson et al., 2012). In oncology clinical research, overall survival is considered the gold standard, this endpoint may not be appropriate as it measures beyond the treatment being assessed (Driscoll & Rixe, 2009; Zhuang, Xiu, & Elsayed, 2009). Researchers have developed their own endpoint that may be appropriate for their research (Collinson et al., 2012), however, the comparison of

efficacy across treatments may become challenging as each endpoint is usually assessed independently. In adaptive studies, the adaptation can be based on a single endpoint or a combination of endpoints (Berry, 2012).

Alternative outcomes have been considered to reduce the duration of trials thus decreasing cost and patient wait time. In oncology, using a surrogate endpoint may decrease the sample size by one third (Fleming, 1996). However, in surrogate endpoint selection, the true effect needs to be assessed while reducing possible noise (Fleming, 1996). The possible overestimation of the effect needs to be considered. Surrogate endpoint selection needs to be considered with the expectation of predicting the true trial outcome using the surrogate outcome. For example, in oncology, treatment response is often used as a surrogate for overall survival. However, colorectal cancer researchers have found through meta-analysis that while there was a tumor response, there was virtually no evidence of improved survival (Fleming, 1996). Other researchers have also reported similar results when comparing the surrogate endpoint to survival (T. T. Chen, Chute, Feigal, Johnson, & Simon, 2000; Villaruz & Socinski, 2013).

Often early phase studies are assessed on single endpoints such as toxicity, in order to identify the maximum tolerated dose (MTD). As a result pharmacodynamics (the study of the effects and mechanisms of drugs) or efficacy are not often considered at all or are not the primary consideration (Ananthakrishnan et al., 2017). Researchers report that in clinical research, consideration of alternative surrogate endpoints related to pharmacodynamics (the study of the effects and mechanisms of drug action) should be examined (Ananthakrishnan et al., 2017). Once again, simulations should be conducted to

assess the operating characteristics including accuracy to identify the MTD using the surrogate endpoint as well as efforts to reduce and acknowledge possible variability in the target population data as well as the endpoint itself (Ananthakrishnan et al., 2017; Fleming, 1996). As stated, consideration of data management logistics should be made when using pharmacokinetic (the study of the movement of the drug in the body) and pharmacodynamics endpoints. Sample collections related to study endpoints, such as pharmacodynamics will need to be conducted on an expedited timeframe to aid in decision making.

Summary and Conclusions

Based on the information presented, researchers appear to have difficulty in adopting innovative methods as they relate to oncology clinical research. Methods such as the 3+3 design continue to be used, even when they have been shown to be ineffective or are used erroneously. Increased knowledge of underlying assumptions related to traditional and adaptive methods is a barrier that needs to be overcome to improve cancer treatment development for the target population. While there may be a high learning curve for adaptive methods, shared knowledge as well as software development and access need to be addressed. The frequency and the operating characteristics through simulations of adaptive versus traditional methods, has been compared in clinical research. The real-world late stage outcome results have not been assessed in relation to the design method used in the early phase studies. The current study assessed the association between early stage design methods (adaptive versus traditional) and the association to late stage outcome results.

Chapter 3: Research Method

Even though adaptive methods have been available since the 1970s with appropriate computational support and knowledge since the 1990s (A. R. Brown et al., 2016), researchers have been reluctant to utilize the design methods. Uptake of adaptive methods, however, has been assertively accepted at some research institutions in comparison to others, likely due to awareness, training, education, feasibility, expertise, computational and programming availability (S.-C. Chow & Chang, 2008; J. A. Quinlan & Krams, 2006; Viele & McGlothlin, 2017). Appropriate communication with the study team to increase comfort, understanding, dissemination and collaboration with regulatory bodies is also critical in the study development process and to increase understanding and design acceptance (Kairalla et al., 2012; Viele & McGlothlin, 2017). Increased awareness, training and expertise need to be improved to allow for increased innovative methods utilization.

I investigated potential associations of early phase study design research outcomes that utilize traditional versus adaptive methods on late stage results in oncology clinical trials. By comparing traditional versus adaptive methods in early phase studies and the downstream late stage clinical trial outcome, the possible benefits related to late stage study success leading to improved patient treatment options, were examined. As drug development and approval are based on cancer type (Barabási et al., 2011), study phase (U. S. Food and Drug Administration, n.d.-b) and treatment classifications (Siddiqui & Rajkumar, 2012), these potential effect modifiers were included in the analysis model. The study examined the potential effect modifiers including phase of

study, treatment classification and type of cancer, on the association between use of adaptive versus traditional methods (e.g. 3+3) in early phase studies and the results outcome of the late stage study. Additional variables such as surrogate endpoint and planned sample size used in the phase III study were assessed in sensitivity analyses.

In this chapter, the research design, methodology and rationale with respect to the study is presented. The operational characteristics are summarized including sample size of the number of studies (the unit of analysis), inclusion and exclusion criteria and sampling procedure of the oncology studies in the Clinical Trials database. The research questions and hypotheses, as well as the related data analysis plan and key analysis variables and operational definitions were also included. Possible threats to validity and ethical procedures related to the Clinical Trials database and Internal Review Board (IRB) research requirements at Walden University, was also addressed.

Research Design and Rationale

This quantitative, nonexperimental, noninterventional, retrospective observational study was analyzed using a Bayesian logistic regression to assess the association between early phase design methods and late stage outcomes in oncology clinical trials. Data for this analysis was extracted from the National Institute of Health Clinical registry and results database (National Institute of Health, 2018). The data is extracted on a daily basis by the Clinical Trials Transformation Initiative (CTTI) Aggregate Analysis of ClinicalTrials.gov (AACT) (Clinical Trials Transformation Initiative, n.d.-a). In addition, researchers also have the option to extract the data independently. Individual studies were the units of the analysis. The Bayesian logistic regression methods were used for the

analysis as asymptotic approximations are not necessary (SAS, n.d.-c) like they are when using Frequentist methods, thus estimates should be more reflective of the data. In addition, when using Frequentist methods, the asymptotic distribution can break down leading to questionable results, which is not the case when Bayesian models are used (Modlin, 2018). Thus, to be expected, researchers have reported that Bayesian techniques have given more accurate results in comparison to classical methods (Guardia-Olmos, 2008; Ogunsakin & Siaka, 2017; Yi, Kaklamani, & Pasche, 2011). The Bayesian analysis parameters and equivalent confidence intervals (credible intervals) also allow probabilities to be utilized and interpreted, a common error when interpreting Frequentist's model estimates (Modlin, 2018). In addition, if a similar analysis related to the use of adaptive studies and their association to late stage outcomes is conducted in the future, the posterior estimates from this model can be used as priors for that analysis based on the Bayesian modeling conducted in this study.

The categorical endpoint used in the analysis was based on the late stage results reported in the Clinical Trials database. If clarification was needed on the late stage results, I reviewed study related publications or study researchers were contacted. The endpoint categories were favorable, equivalent, or nonfavorable. The favorable category was utilized if the late stage clinical trials results reported in the database were statistically or clinically significant for the experimental treatment in comparison to the standard of care. The endpoint was classified as nonfavorable if the late stage clinical trials results were reported where the standard of care was reported to be statistically or clinically significantly better in comparison to the experimental treatment. For

bioequivalence or noninferiority studies and no significant endpoint differences were found, the study was coded as equivalent (favorable). A similar outcome result classification system was utilized by the researchers Rasmussen, Lee and Bero (2009).

While the clinical trials results were reported in the database, the endpoint categories were identified where possible through statistical programming. The confirmation of the identified categories were manually reviewed by me, and a percentage were verified through the database, publication review, or researcher contact. Similarly, the covariate classifying the early stage design methods (adaptive versus traditional) was also extracted from the database, if available in the database study description. If the clinical trial design was not included in the database description, the design classification variable was captured via publications, or researcher contact. Once again, confirmation of the categories was manually reviewed and a percentage was verified through publication review as well as researcher confirmation as needed. The remaining variables included in the analysis were included in the database and are described in Table 4 below.

The analysis model was:

Late phase study results (favorable/unfavorable) = early phase design (adaptive/traditional) + early phase + experimental treatment classification + type of cancer + drug classification + duration between early and late phase studies (months) + study funding

This study could aid in understanding the influence of adaptive design on cancer treatment development. In addition, this study could increase knowledge related to

adaptive design, which has been identified as being a barrier related to the adoption of the innovative methods (Kairalla et al., 2012).

Methodology

Population

Aggregated study data reported in the publicly available Clinical Trials database were extracted in March 2018 from the related results database using the Clinical Trials Transformation Initiative (CTTI) Aggregate Analysis of ClinicalTrials.gov (AACT) (Clinical Trials Transformation Initiative, n.d.-a; National Institute of Health, 2018). The database is a cloud-based resource used by researchers, clinicians and patients to find information related to clinical studies for specific diseases and conditions (National Institute of Health, 2018). The registry of studies was established for all funded studies and was the result of the Food and Drug Administration Modernization Act of 1997 (Food and Drug Administration, 1997). The database is a requirement of the specified legislation to be established and maintained by the United States Department of Health and Human Services (HHS). Studies for the treatment serious or life-threatening diseases or conditions are required to register in the database. Further the results database including study outcomes and adverse events were made available starting in 2008. The extracted data were aggregated by study which was the unit of analysis. No additional access or permission was needed to use the database information, as the data is publicly available and accessible.

Sampling and Sampling Procedures

Stratified random sampling based on cancer type, treatment and drug classification was conducted to allow for appropriate sample size in order to increase possible power for comparisons in sub-populations. Within strata, clinical trial order was random and identified as adaptive or traditional. Random numbers were assigned to each study and ordered from lowest to highest. Studies were included in the sample in that order until the sample size was met within each design method. Table 3 provides an example of studies within cancer type and treatment type.

Sampling Strategy

Stratification variables were used to ensure that analysis within each cancer type, treatment classification, experimental drug classification and funding source could be conducted. Researchers have noted that within the Clinical Trials database, approximately 37% (158/428) to (Hatfield et al., 2016) 42% (142/336) (Bothwell et al., 2018) of the Phase II to Phase III studies used adaptive methods.

Table 3

Random Study Order Identification Within Strata

Cancer Type			Treatment type	
Cancer type 1	Cancer type 2	Cancer type 3	Drug	Radiation
NCT_ XXXX1	NCT_ XXXX2		NCT_ XXXC1	
NCT_ XXXX3	NCT_ XXXX4		NCT_ XXXD1	
NCT_ XXXX7	NCT_ XXXX9	…	NCT_ XXXC3	

Inclusion Criteria

Only studies for oncology intervention treatment were included. Intervention treatments include drug, biological, surgery, radiation and any combination. Combination treatments were also be eligible for the analysis and appropriately identified in the analysis model. Studies initiated between, January 1, 2000 to December 31, 2016 were included in the analysis. The latter cutoff date allowed enough time (approximately one to three years) for the studies to mature and results to be included in the database.

Exclusion Criteria

Early phase studies for healthy volunteers were excluded, as the studies cannot be considered interventional. Oncology related palliative care or noninterventional studies were excluded in the analysis.

Archival Data

Data from the National Institute of Health Clinical Trials registry and results database were used for this study (National Institute of Health, 2017a). The data was extracted by the Clinical Trials Transformation Initiative on a daily basis and available using the recommended software (Clinical Trials Transformation Initiative, n.d.-a, n.d.-b). In 1997, the Food and Drug Modernization Act (FDAMA) was enacted requiring a registry for efficacy trials for serious and life-threatening conditions, including oncology studies (National Institute of Health, 2015b). The law was amended in 2007, (HR 3580, the FDA Amendments Act of 2007) where a legal requirement for the registration of the trials of drugs was implemented (National Institute of Health, 2017b). The clinical trials results database was initiated in 2008.

Researchers have reported that the average duration of Phase III oncology clinical trials is between 1 and 3 years (Pregelj, Verreynne, & Hine, 2015). Therefore, studies initiated in 2016 may be close to reaching maturity of the primary endpoint. Thus, oncology studies initiated prior to 2016 (up to December 31, 2015) were included in the analysis.

In 2004, the International Committee of Medical Journal Editors required researchers to register in ClinicalTrials.gov or an equivalent registry prior to the first patient to be enrolled in their study (International Committee of Medical Journal Editors, 2017). This criterion had to be met in order for the publication to be considered for acceptance into their journal. This committee is represented by most major journal publications; thus, there is increased motivation for the researchers to register their studies not just in the United States, but also across the globe. As a result of the FDA enactments as well as the International Committee of Medical Journal Editors, the clinical trials database should be representative of clinical trials being conducted in the United States and possibly around the world. Further, effective in January 2015, the National Cancer Institute (NCI) required that all NCI-supported intervention study results need to be reported publicly within 12 months of the study completion date (National Institute of Health, 2015a), thus improving cancer clinical trials results representation in the database.

Instrumentation and Operationalization of Constructs

Independent Variable

The categories of the study design were based on the early phase study design specifically adaptive or traditional (see Table 4 for definitions). As a result of the changing definition of adaptive studies, the ambiguous definition evolving over the years (Bothwell et al., 2018; Christopher S Coffey et al., 2012; Kairalla et al., 2012) and the self-reporting of study design methods categorization (Bothwell et al., 2018), I reviewed the database and related publications to provide clarity of the study description design. Adaptive studies were defined to be clinical study designs that use accumulated information or data of the study to modify aspects of the study as it continues (Christopher S Coffey et al., 2012). The adaptations in these designs are predefined. This definition is consistent with the Pharmaceutical Research and Manufacturers of America (PhRMA) Working Group (Gallo et al., 2006). Traditional studies were defined to be studies that did not satisfy the definition for adaptive. Adaptive methods that were defined to be less well-understood were designs that have not been frequently utilized thus there were less experiences with the methods (U. S. Food and Drug Administration, 2010). Each study within the Clinical Trials database has a unique identification associated, thus confirming classification with publications and facilitating communication with researchers on specific studies is simplified. Classification of design methods was based on the Pharmaceutical Research and Manufacturers of American (PhRMA) Working Group definition (Gallo et al., 2006). Subdivision of adaptive designs

into well-understood versus less well understood was also be used based of FDA definitions (U. S. Food and Drug Administration, 2010).

Operationalization of Variables

Table 4 provides the variable information that was used for the analysis. Included in the table are the variable definitions, the format of the variables within the Clinical Trials database, as well as the value of the variables for the analyses.

Table 4

Analysis Model Variables, Description and Possible Values

Variable	Definition	CT.gov format	Values
Dependent variable			
Outcome of late-phase study	Late-stage study results. Results are based on the original study operation characteristics and statistical significance. The statistical outcome of the late-phase study (favorable = experimental treatment statistically better than comparison; equivalent = experiment and comparison are statistically equivalent; nonfavorable = experimental treatment is found to be better than the comparison treatment). For studies that are not driven by statistical significance, clinical relevance (effect size) will determine the outcome classification.	Free text, categorized for variable values.	Endpoint: favorable, equivalent, unfavorable
Independent variables			
Design classification	Adaptive studies are defined to be clinical study designs that use accumulated information or data of the study to modify aspects of the study as it continues (Christopher S Coffey et al., 2012). The adaptations in these designs are predefined. Traditional studies will be defined to be nonadaptive studies.	Free text	Adaptive, Traditional
Early phase	Phase of early-stage study	Same as value	Categorical: 0, 1a, 1b, 1, 2, 2a, 2b

(continued)

Variable	Definition	CT.gov format	Values
Experimental treatment classification	Interventional treatment classification	Categorical	Drug Device Biological/vaccine, Procedure/Surgery Radiation Combination Other
Cancer type	The cancer site or type that is being researched in the early- and late-stage study.	Same as reported value	Breast, lung, pancreatic, bone etc.
Experimental drug classification	WHO drug classification	Categorical variable	EGFR or VEGF inhibitors, monoclonal antibody, proteasome inhibitor, immunotherapy etc.
Duration between studies (months)	Time between the end of the early-stage study and the beginning of the late-stage study	Dates; will take difference for duration	Time in months
Study funding source	Who is sponsoring the early stage study?	Reported funding (grant, other)	Private or public funding
Sensitivity analysis variables			
Type of endpoint	Type of endpoint used. Any endpoint that is not overall survival such as progression free survival or objective response will be considered a surrogate.	Free text	Surrogate, clinical or both
Sample size	The number of patients planned to be enrolled in the early-phase studies.	Numeric	Numeric
Biomarker	Did the study include a biomarker as an endpoint in the study or not?	Categorical	Binomial (yes, no)
Adaptive classification	For those studies that are adaptive, are they are classified as well-understood or less well-understood designs as defined by the FDA (2010).	Categorical	Well-understood, less well-understood, not adaptive

Data Analysis Plan

All analyses were performed using SAS 9.4 or higher (SAS, n.d.-b). Data was downloaded from the CTTI AACT relational database, which included the study characteristics such as phase, type of intervention, population, sample size, among other variables in aggregate form. Target analysis variables were extracted, including study outcome, phase, treatment, cancer type, experimental drug classification and study related dates. CT.gov studies included in this analysis were assessed and reported for completeness.

When studies are entered into the Clinical Trials database, records are reviewed for accuracy and content by the NIH database administrators (National Institute of Health, 2010). The database administrators also conduct programmatic and manual internal consistencies with respect to other study related information associated with the clinical trial (Zarin, Tse, Williams, Califf, & Ide, 2011). Once the results of the study have been reported, consistency within the record are assessed, as well as comparing information that is relevant to the appropriate field (National Institute of Health, 2009). Protocol and results data entry guidance is also provided by the database administrators to the researchers which should improve data quality (Tse, Williams, & Zarin, 2009; Zarin et al., 2011).

For any clarification related to early phase design selection, late stage outcome results, study related researchers were contacted and/or publications were reviewed using the Clinical Trials unique study identifier for the search. Previous researchers (Hatfield et al., 2016) have also classified study designs (adaptive classifications) using the Clinical

Trials database. The Hatfield et al. (2016) analysis datasets, including study design classification variable, were publically available and were used for study design verification in this analysis. Bothwell et al. (2018) also conducted a study of adaptive study classifications using the Clinical Trials database. While the data for the Bothwell analysis was not publicly available, the researchers shared a list of adaptive classified studies that they included in their analysis. This data were also be used for validation of study classification in this study.

In addition, an independent review and classification of 10% of the studies were conducted. A kappa statistic was used to measure the agreement between the independent data reviewers. If the discrepancy rate was greater than 15%, an additional 10% of study design classifications were reviewed. Prior to the analysis of this study, the two independent reviews were reconciled.

Sample Size and Power Estimation

As previously noted, the failure rates of oncology treatments are as high as 66% (Berry, 2011). Using these estimates, assuming a positive outcome rate of 37% for late stage studies where adaptive methods (nonadaptive: 22%) were used, there is 85% power with a sample size of 425 studies (adaptive=125; nonadaptive=300). Under these assumptions, as well as a polynomial endpoint (phase III study outcome: positive, equivalent, negative) and early phase study design method (adaptive or not), a two-sided alpha of 0.05 was used to estimate the sample size. Sample size estimates were computed using Nquery Advisor 8.0 (Statistics Solutions, n.d.).Table 5 provides power estimates if

the proportion of positive results within the traditional versus adaptive design methods was varied.

Table 5

Sample Size Under Assumptions of Proportion, Power and Type I Error for Target Population

Power (%)	Type I	Proportion		Sample Size	
		Traditional	Adaptive	Traditional	Adaptive
85	0.05	0.22	0.37	300	125
75	0.05	0.22	0.35	300	125
65	0.05	0.25	0.37	300	125
53	0.05	0.22	0.32	300	125
48	0.05	0.27	0.37	300	125

To allow for increased power to compare cancer types, treatment and drug classifications, a stratified identification of studies to be included in the analysis was used. The order of the possible studies included within each stratification variables was random. Clinical trials were included in the analysis in the randomization order until the target sample size was met.

Table 6 provides additional power estimates for sub-populations under a variety of assumptions for power and proportion of positive results for late stage studies when traditional versus adaptive methods were used for the early stage studies. For example, assuming a two-sided type I error (the probability of falsely rejecting the null hypothesis) of 0.05 and 22% of positive outcomes of late stage studies where traditional methods were used and 46% where adaptive methods were used, there is 80% power if 150 studies (traditional=100, adaptive=50) were identified in the sub-population. In addition, if the

different rates of adaptive versus traditional were identified, alternate power estimates were provided.

Table 6

Sample Size Under Assumptions of Proportions of Positive Outcomes, Power and Type II Error for Subpopulation

Power (%)	Type I	Proportion positive outcome		Sample size	
		Traditional	Adaptive	Traditional	Adaptive
80	0.05	0.22	0.46	100	50
60	0.05	0.30	0.50	100	50
42	0.05	0.25	0.37	100	50
81	0.05	0.22	0.32	75	30
28	0.05	0.22	0.37	75	30

Research Questions and Hypotheses

RQ1: To what extent is there an association between design methods used for early phase oncology studies (adaptive versus traditional) and the outcome of late stage clinical trials?

H_o1: There is no association between early phase oncology studies (adaptive versus traditional) and the late stage clinical trial outcomes.

H_a1: There is an association between early phase oncology studies (adaptive versus traditional) and the late stage clinical trial outcomes.

RQ2: In specific cancer types (e.g. breast cancer, lung cancer etc.), what is the association between the design methods used for early phase oncology studies (adaptive versus traditional) and the outcome of the late stage clinical trials?

H_o2: The association between early phase oncology design methods (adaptive versus traditional) and the outcome of the late stage clinical trials does not differ between specific cancer types.

H_a2: The association between early phase oncology design methods (adaptive versus traditional) and the outcome of the late stage clinical trials differs between specific cancer types.

RQ3: How does the treatment classification modify the relationship between the design methods used for early phase oncology studies (adaptive versus traditional) and the outcome of the late stage clinical trials?

H_o3: The association between early phase oncology design methods (adaptive versus traditional) and the outcome of the late stage clinical trials does not differ between specific treatment classification (e.g. surgical, adjuvant, radiation, etc.).

H_a3: The association between early phase oncology design methods (adaptive versus traditional) and the outcome of the late stage clinical trials does differ between specific treatment classification (e.g. surgical, adjuvant, radiation, etc.).

Analysis Plan

Analysis Model

A multivariable Bayesian logistical regression model was used for this study. The comparator covariate was the study identified design methods (adaptive versus traditional) used by the early phase studies. The outcome (or response) variable was the late stage outcome that had three categories (favorable, equivalent and nonfavorable). Favorable is when the late stage study outcome results were statistically significant or

clinically significant per study design and in favor of the experimental treatment. A study was deemed to be nonfavorable if the results did not reach statistical or clinical significance in favor of the experimental treatment. For bioequivalence or noninferiority studies and no significant endpoint differences were found as designed, the study was coded as equivalent (favorable). A similar classification system was utilized by the researchers, Rasmussen et al. (2009).

Covariates included in the model were phase of study, treatment classification, cancer type, experimental drug classification (if applicable), duration between studies and study funding. As researchers have found that there is no single treatment for all cancers, the treatment classification, cancer type and experimental drug classification should be included in the model (Barker et al., 2009; Carey et al., 2010; Kim et al., 2012; Lonial & Nooka, 2016). Phase of treatment was included as it is unclear the impact of each development phase of study (phase 0, 1, 2) and the contribution of each phase with respect to the outcome of the late phase study results.

Duration between the early phase study and the late stage study was also included as the shorter or longer lag time, were likely to have less influence on an ongoing study. As such a random forest analysis of the time between the early and late stage studies and outcome was used for categorical classification. The classification of the duration of time variable was included in the analysis model. The duration of time classification variable was not expected to be linear. If no significant classifications were identified using the random forest methods, the duration of response would be classified by quartiles. Study

funding, specifically private versus public, has been found to be associated with adaptive method design versus not (Kairalla et al., 2012).

Frequentist Modeling Techniques

Standard logistic regression analysis for univariate modeling with a logit link function and binomial distribution was used for univariate analyses. PROC LOGISTIC was used for the logistic modeling. Odds ratios and 95% confidence intervals were computed and presented. Type 3 p values will be used to assess the statistical significance of all variables including those classified as categorical. Type 3 p values assess the overall influence of the variable including all the categories within the variable as well as interaction terms, if applicable. For forward and backward stepwise modeling, HPGENSELECT was used with a binomial distribution and a logit link. Akaike Information Criteria corrected for bias (AICC) was used to determine which variable should be removed from the model.

Bayesian Modeling Techniques

A Bayesian logistic regression analysis for modeling with a logit link function and binomial distribution was used. PROC GENMOD was used for the logistic modeling with contrast statements of linear combinations of parameters included in the model, including interaction terms where specified. PROC PLM was used for post processing estimations of model estimates including the odds ratios and the corresponding credible intervals. The credible interval is an interval that the true but unobserved parameter falls within a specific probability based on quantiles (Stokes, Chen and Gunes, 2014). The 95% highest posterior density (HPD), the minimum interval, was computed. The analysis

was conducted with outcome of the late stage study (favorable, equivalent and nonfavorable). The noninformative prior distribution with mean of zero and a large variance (1x10^6) for the independent parameters in the model was assumed to be multivariate normal. The burn-in period for the Markov Chain Monte Carlo was 2000 samples with 20,000 simulations, which is the default setting in SAS and conventionally accepted as initial frequencies (Stokes, 2014). However, adjustments to the burn-in period or number of simulations were made based on the convergence and auto-correlation model assessment statistics. The Markov chain techniques were used for sampling using the Gamerman algorithm (Gamerman, 1997).

Assessment of the Model

Convergence was assessed on all parameters using the trace plots. The trace should show good variability across the plot for good mixing and low autocorrelation across Markov chain samples. In addition, the posterior autocorrelation was assessed to determine if convergence has been attained. The burn-in period was extended if good mixing was not evident in trace plots through an equilibrium distribution or if convergence was not been attained. Convergence diagnostic tests was also utilized including Gelman-Rubin (1992), Geweke (1992), Heildelberger-Welch (1983) and, Raftery-Lewis (A. E. Raftery & Lewis, 1996; A. E. L. Raftery, S. M., 1992). Table 7 below provides the description and measures of each convergence test.

The Deviance Information Criterion (DIC) (Spiegelhalter, Best, Carlin, & van der Linde, 2003) is a Bayesian generalization of the Akaike information criteria (Akaike, 2011) and was used to assess the overall model as well as the contribution of each

variable. Dispersion and heterogeneity were assessed and if needed, the covariance matrix was scaled. Table 7 provides details on convergence tests that were used to assess the model.

Table 7

Bayesian Model Convergence Tests, Description and Purpose

Convergence test	Description and purpose
Gelman-Rubin	Multiple Markov chains are assessed to determine if they converge to the same target distribution or not. If the test fails, a longer burn in period may be needed. A large test value indicates the null should be rejected.
Geweke	Compares means from early and late parts of the Markov chain to see whether convergence has been attained (Ho: $\mu_1=\mu_2=\ldots.=\mu_j$)Assess whether the mean estimate of the Markov chain is stable over time. A large Z statistic (or small *p*-value) indicates rejection of the null.
Heidelberger-Welch	Consists of a two part test. The first part assesses the stationarity of the Markov chain (Ho: Chain comes from the covariance stationary process). The second part of the test is the half-width best checks where the Markov change is assessed as to whether or not the sample size is adequate to estimate the mean value accurately. This one sided test is rejected if the *p*-value is greater than 1-a. Also need to specify epsilon, but not sure where this fits into the test.
Raftery-Lewis	The posterior percentiles was used to summarize the parameter estimates, the R-L diagnostics assessed the accuracy of the estimated percentiles. The closer the ratio of sampled values of a parameter versus the sample number is to 1, the less correlated the samples are. Note that this test focuses on the percentiles and does not assess the convergence overall. This assessment will aid in determining whether the burn in period is appropriately sized or not. This test tends to be conservative in that, the statistic will suggest more iterations than necessarily needed.
Effective Sample Size (ESS)	ESS=n/tau where n is the total sample size and the autocorrelation time (tau) is basically the sum of the autocorrelation. A low ESS or high tau indicates a bad mixing of the Markov chain. This ESS assessment is an indicator of correlation between successive samples. The higher the correlation between successive samples, the less information gained.

(SAS, n.d.-a; Stokes, 2014)

Model Interpretation

Descriptive statistics including the mean, median and range for continuous variables were computed. For categorical variables, frequencies were computed. PROC GENMOD was utilized where the coefficients of each covariate was estimated, adjusted for the other variables in the model. The odds ratio along with the associated 95% credible interval will be computed. The posterior summaries and credible intervals were used to assess parameter significance. If the odds ratio was greater than one, then there were higher odds of a positive outcome with respect to the exposure (i.e. adaptive methods). Similarly, if the odds ratio is less than one, there are lower odds of a positive outcome with respect to the exposure. The expectation is that when adaptive methods are used, there will be higher odds of a positive late phase outcome. If the 95% credible intervals include the value one, then the results are not statistically significant.

Multiple Comparisons

Three hypotheses were evaluated, however, adjustments type I errors for multiple comparisons was not be employed. In order to test each hypothesis, the full model was assessed for significance along with the reduced parameter model (parameter of interest + design methods). The first hypothesis was to compare whether the early phase design method is related to the outcome of the late phase study. The statistical significance of this hypothesis was assessed using the credible intervals (equivalent to classical confidence intervals) in the logistic model. The Deviance Information Criterion (DIC) was compared across the full versus the reduced parameter model to assess variable contribution. For the second and third hypothesis, a similar test was conducted within

each cancer type as well as within each treatment classification. Subgroup analyses was conducted and assessed regardless of statistical significance.

Threats to Validity

For any analysis there can be threats to validity that should be acknowledged and countered to the extent possible. Often when multiple studies are combined, publication bias may be a concern. Publication bias is the result of more positive studies being published in journals, in comparison to results with negative studies. However, due to the FDAMA and FDAAA requirements of registry for efficacy trials for serious and life-threatening conditions including oncology studies (National Institute of Health, 2015b, 2017b), the Clinical Trials database should be representative of oncology clinical research occurring in the United States. The registry requirement for publication in major journals prior to patient enrollment (International Committee of Medical Journal Editors, 2017) also improves the quality and representation of the database.

Researchers have shown that the Clinical Trials database is of high quality with respect to publications as reported by Hartung et al. (2014). An independent review of a randomly extracted sample was conducted to assess discrepancies between the results reported in the Clinical Trials database and the related publications. Of the 110 trials that Hartung et al. (2014) randomly selected from the database, approximately 15% reported primary outcome descriptions and 20% reported primary outcome value inconsistencies. Any possible discrepancies identified for the current analysis were reported to the database administrators.

While the Clinical Trials database has a requirement that the results from studies be entered within one year of maturity for designated studies, researchers have noted that delays in reporting appears to be a consistent problem (Anderson et al., 2015). Increasing the time tolerance for reporting did improve the reporting percentage, however despite ethical and legal obligations as required by United States law, delays in reporting is prevalent (National Institute of Health, 2017a). Missingness within included studies were reported and assessed for possible bias and patterns.

In addition, as noted by previous researchers, with the ambiguity of the definition of the classification of adaptive studies and self-reporting of classification (Bothwell et al., 2018) may negatively impact the results of the current study. However, as noted previously, 10% of the studies included in this analysis were independently verified and compared. If the discrepancy rate was noted to be greater than 15%, an additional 10% of study design classifications were reviewed.

Ethical Procedures

Ethical issues can arise during enrollment as well as data collection of any clinical trial (Creswell & Creswell, 2014). As this study uses archival data, ethical related issues through Internal Review Boards (IRB) or Central Review Boards (CRB) were addressed within each study and enrolling site, prior to the initiation of this study due to human subjects being enrolled and treated for their cancer (National Institute of Health, n.d.). Prior to enrolling in the Clinical Trials database, the researchers need to get review board approval before the Clinical Trials database status is changed to recruiting.

The data used for this analysis, collected between the years 2000 and 2016 in the Clinical Trials database is aggregated, thus there were no concerns for patient data identification, as the data includes no patient identifiers. No individual patient information is available in the database. Prior to my study initiation and data analysis, I obtained an IRB approval through Walden University. The data for the analysis is publicly available and can be downloaded directly from the Clinical Trials database website without permission.

Summary

This study was conducted to assess the relationship between early phase study design methods and the outcomes of late stage study results. The quantitative nonexperimental noninterventional retrospective observational study used a Bayesian logistic regression analysis using data from the Clinical Trials database. The unit of analysis was individual studies where the data is aggregated within the database. Additional covariates included in the model included experimental treatment classification, type of cancer, drug classification, study funding as well as the duration between the early and late phase study. This study could aid in understanding the influence of adaptive design on cancer treatment development and may also increase knowledge related to these innovative methods in cancer research.

Chapter 4: Results

In this quantitative observational study, I investigated the potential association between early phase study design research that utilized traditional versus adaptive methods and late stage results in oncology clinical trials. By comparing traditional versus adaptive methods and the downstream clinical trial outcome, the association with late stage study outcome was assessed. As drug development and approval are based on cancer type, study phase and treatment classifications, these potential effect modifiers were included in the analysis model (Barabási et al., 2011; U. S. Food and Drug Administration, n.d.-b; Siddiqui & Rajkumar, 2012), to determine whether the association between use of adaptive versus traditional methods (i.e. 3+3) in early phase studies and the results outcome of the late stage study, is different in subgroups defined according to levels of these factors.

The hypotheses that were tested include:

H_o1: There is no association between early phase oncology studies (adaptive versus traditional) and the late stage clinical trial outcomes.

H_a1: There is an association between early phase oncology studies (adaptive versus traditional) and the late stage clinical trial outcomes.

H_o2: The association between early phase oncology design methods (adaptive versus traditional) and the outcome of the late stage clinical trials does not differ between specific cancer types.

H_a2: The association between early phase oncology design methods (adaptive versus traditional) and the outcome of the late stage clinical trials differs between specific cancer types.

H_o3: The association between early phase oncology design methods (adaptive versus traditional) and the outcome of the late stage clinical trials does not differ within specific treatment classification (e.g. surgical, adjuvant, radiation, etc.).

H_a3: The association between early phase oncology design methods (adaptive versus traditional) and the outcome of the late stage clinical trials does differ within specific treatment classification (e.g. surgical, adjuvant, radiation, etc.).

The purpose of this chapter is to provide a description of the sample of data obtained from the ClinicalTrials.gov clinical trials database for oncology studies as well as provided a description of the results of the statistical analysis conducted. The results are described using tables and figures. Narratives are also provided to support assessments of the statistical significance of planned and sensitivity analyses. A summary of the findings with respect to the research questions is also provided.

Source Data

This retrospective study was an analysis of archival database including oncology clinical trials maintained by the National Institute of Health and aggregated by the Clinical Trials Transformation Initiative (CTTI) (Clinical Trials Transformation Initiative, n.d.-a; National Institute of Health, 2018). Summary variables were also added to the database by CTTI, such as number of treatment arms, device study, radiation, to name a few. The oncology studies included in the analysis were interventional for

patients with disease but excluding studies that accrued healthy volunteers that were initiated after November 1, 1999 to December 31, 2016. For studies included in the analysis, the results of the primary endpoint had to be reported in the database by the investigators at extraction. The data is publicly available and was extracted from the Clinical Trials database on March 26, 2018, which was prior to IRB approval. As the database is actively accruing data and no codebook with descriptive data were available, the first extract was planned to be used to identify variables of interest and get a better understanding of the database structure. The second extraction after IRB approval was planned however, due to challenges identified with the first extraction with the size of the database, hidden characters in the extracted datasets creating errors, the need for multiple database administrators' assistance and no available codebook with summary statistics as the database is active, the original extraction was used for the analysis. However, as the data is archival, publicly available for extraction and aggregated with no individual patient identifiers, there continues to be no risk to patients.

Data Analysis

At data extraction there were 269,310 studies of which 3,040 oncology intervention studies that did not include healthy volunteers with the indicator that results were available. A summary of the Clinical trials database is found in Table 9. Oncology intervention clinical trials were included in this study. Studies were sorted by covariates (disease condition, treatment type and year initiated) and randomly assigned a number based on these covariates. Increased weights were assigned to more recent research as researchers have noted that adaptive methods have increased in use over the years

(Hatfield et al., 2016). Increasing the weights of recent studies, increases the probability of adaptive studies to be included in this study. Studies included in this research were randomly added and a sample size of 381 was attained (286 early phase, 95 late phase).

Of the early phase studies, 220 (76.9%) were deemed traditional design and 66 (23.1%) were classified as adaptive studies. Similarly, of the late phase studies, 56 (58.9%) and 39 (41.0%) were classified as traditional and adaptive design respectively. The overall percentage of adaptive studies was 27.5% ($N = 105$).

Studies that included the most frequent diseases, specifically breast, lung, pancreatic, sarcoma and multiple cancers, were included in the study to try to ensure frequencies of cancers were large enough for primary and sub-study comparisons. However, as previously noted, programmatic filtering of disease condition was not 100% accurate in identifying the correct disease type. In addition, studies that recruited multiple populations were included in the analysis, thus additional cancer types were included in the analysis dataset. Nonetheless, due to the random selection of studies, the sample should be representative of all oncology studies in the database. Table 8 includes a description of the database and studies included in the analysis. Table 9 includes a summary of the reason why studies were excluded from this analysis of which the primary reason was due to accrual problems (32.7%) leading to incomplete results. However the unreported reason for the incomplete results was also frequently noted (29.5%). Figure 3 includes a visualization of the studies in the database, reasons included and excluded as well as a description of the included studies.

Table 8

Summary of Studies Within the Clinical Trials Database

Total	N, %
Number of studies	269,310
Number of studies excluded[a]	203,522
Oncology intervention studies	3819
With results	3040
Random sample	537
Excluded[a]	156
Included in analysis	381
Phase (adaptive, N)	
Early	286 (66)
Late	95 (39)
Design (%)	%
Adaptive	105
Traditional	276

a. See Table 9 for exclusion reasons.

Table 9

Reason Studies Were Excluded from the Analysis

Exclusion reason	n (%)
Number of studies excluded	156
Results outcome	
Incomplete	122 (78.2)
Unknown	18 (11.5)
Unknown design	4 (2.6)
Reason unplanned study stoppage	
Accrual problems	51 (32.7)
Administrative reasons	1 (0.6)
Business decision	13 (8.3)
Complete	2 (1.3)
Drug availability	2 (1.3)
Funding problem	7 (4.5)
Futility or safety concerns	16 (10.3)
Other reason	4 (2.6)
PI logistics	14 (9.0)
Unknown – not reported	46 (29.5)

Interrater Agreement

Five hundred and thirty-seven studies were randomly selected from the clinical trials database. Ten percent (54 studies) of the sample was randomly selected to allow for an independent review of the study outcome and study design. The independent reviewer's classifications were compared to the analysis dataset using the Kappa coefficients, which measures inter-rater agreement, are presented in Table 10. The inter-rater agreement was greater than 94% for each variable with a Kappa coefficient of 0.9206 and 0.8845 respectively indicating a good agreement across raters thus indicating that manual classifications included in the analysis were relatively accurate. As such, independent review of an additional 10% of the sample was not needed.

Table 10

Interrater Agreement for Study Design and Study Outcome Classification

Study design classification	n (%)
N	54
Agreement	96.3%
Conflict	3.7%
Kappa coefficient (95% CI)	0.9206 (0.8126, 1.0000)
Standard error	0.0551
Two sided p value	<0.0001
Study outcome classification N (%)	
N	54
Agreement	94.4%
Conflict	5.56%
Kappa coefficient (95% CI)	0.8845 (0.7592, 1.0000)
Standard error	0.0639
Two sided p value	<0.0001

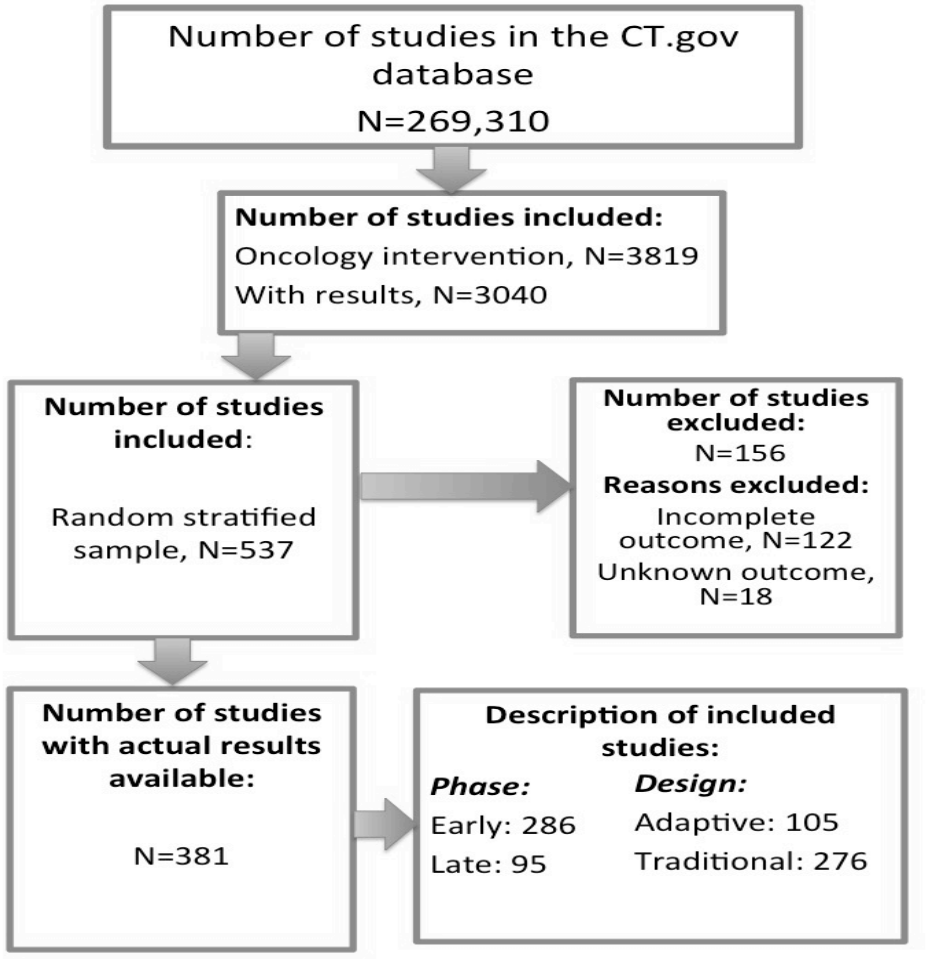

Figure 3. Exclusion and inclusion criteria for studies selected for analysis.

Aggregate Participant Characteristics Within Studies

Table 11 provides descriptive characteristics aggregated for participants within each study included in this analysis. The median sample size for early phase and late phase studies was 22 (Range: 1 to 343) and 158 (8 to 2091) patients respectively. Most early phase studies included only one participating country, primarily the United States (55.6%), whereas late phase studies had a median of 14 participating countries. Most

studies in either phase included adults (age greater than or equal to 18 years). The most common cancers included in early stage studies were lung (12.9%), colorectal (12.9%), breast neoplasms (12.3%) and multiple myeloma (12.6%). Multiple cancers were included in 8.0 percent of early phase studies. Similarly, the most common cancers included in the late stage studies were breast (25.7%), lung (19.9%) and multiple myeloma (12.3%), of which only 1.1% of studies included multiple disease conditions. Most studies in either stage category included both men and women (76.0% and 65.3%, respectively) with age greater than or equal to 18 (Early: 88.1%; Late 91.6%). Additional details of participant characteristics within studies can be found in Table 11.

Table 11

Summary of Participant Characteristics Within Studies (Analysis Units)[a]

		Early phase [b]	Late phase [b]	Total
Number of studies, N (%)		286 (75.1)	95 (24.9)	382
Number of participants (actual)	Median (min, max, STD, N)	22 (1, 343, 47.9, 284)	158 (8, 2091, 275.5, 94)	36 (1, 2091, 167.5, 378)
Number of participating countries	Median (min, max, STD, N)	1 (1, 30, 4.7, 278)	14 (1, 62, 12.4, 92)	1 (1, 62, 8.9, 370)
Minimum age of study participants[c] (years) (n, %)				
	≥ 2 years	1 (0.4)	0	1 (0.3)
	≥ 15	1 (0.4)	0	1 (0.3)
	≥ 16	2 (0.7)	1 (1.1)	3 (0.8)
	≥ 18	252 (88.1)	87 (91.6)	339 (89.0)
	≥ 19	3 (1.1)	0	3 (0.8)
	≥ 20	18 (6.3)	4 (4.2)	22 (5.8)
	≥ 21	2 (0.7)	2 (2.1)	4 (1.1)
	≥ 25	1 (0.4)	0	1 (0.3)
	≥ 30	1 (0.4)	0	1 (0.3)
	≥ 40	1 (0.4)	1 (1.1)	2 (0.5)
	≥ 45	1 (0.4)	0	1 (0.3)
	≥ 60	1 (0.4)	0	1 (0.3)
	≥ 65	2 (0.7)	0	2 (0.5)
Maximum age of study participants[c] (years) (n, %)				
	≤ 18 only	1 (0.4)	0	1 (0.3)
	≥ 65	38 (13.3)	14 (14.7)	52 (13.7)
	Not specified	247 (86.4)	81 (85.3)	328 (88.1)
Cancer type[d], N (%)				
	Breast neoplasms	79 (12.3)	44 (25.7)	123 (15.1)
	Lung neoplasms	83 (12.9)	34 (19.9)	117 (14.4)
	Multiple myeloma	81 (12.6)	21 (12.3)	102 (12.5)
	Colorectal Neoplasms	83 (12.9)	15 (8.8)	98 (12.0)
	Prostatic neoplasms	58 (9.0)	19 (11.1)	77 (9.4)

(continued)

		Early phase [b]	Late phase [b]	Total
Cancer type[d], N (%) Continued				
	Pancreatic Neoplasms	38 (5.9)	10 (5.8)	48 (5.9)
	Ovarian neoplasms	30 (4.7)	8 (4.7)	38 (4.7)
	Head and neck Neoplasms	24 (3.7)	1 (0.6)	25 (3.1)
	Stomach neoplasms	15 (2.3)	3 (1.8)	18 (2.2)
	Neoplasms, solid	14 (2.2)	0	14 (1.7)
	Anal neoplasms	11 (1.7)	0	11 (1.3)
	Gastrointestinal Neoplasms	6 (0.9)	4 (2.3)	10 (1.2)
	Other[c]	122 (18.9)	12 (4.4)	134 (16.4)
	Multiple conditions	23 (8.0)	1 (1.1)	24 (6.3)
Genders included in study, N (%)				
	Males only	34 (11.9)	11 (11.6)	45 (11.8)
	Females only	33 (11.5)	20 (21.1)	53 (13.9)
	All	218 (76.0)	62 (65.3)	280 (73.5)
	Not reported	1 (0.4)	2 (2.1)	3 (0.8)
Region enrolled				
	Ex-United States only	52 (18.2)	35 (36.8)	87 (22.8)
	United States only	159 (55.6)	3 (3.2)	162 (42.5)
	U.S. and ex-U.S.	67 (23.4)	54 (56.8)	121 (31.8)
	Not reported	8 (2.8)	3 (3.2)	11 (2.9)

a. Participant data within studies was aggregated.
b. Study phase of less than III as well as seamless studies of Phase II/III were considered early.
c. Maximum age was often not specified indicating that adults of any age could be included in the study. Some studies reported extreme age values such as 105, rather than leaving the field blank, still indicating that adults of any age could enroll in the study.
d. A complete list of disease conditions can be found in Appendix XX. Overall counts within a classification of less than 10 were combined into the 'Other' category.

Characteristics of the Included Studies

As a result of the criteria to be included in this study, most studies were reported by researchers to be interventional (Early: 99.7%; Late: 97.9%) of which most were drug treatment studies (Early: 96.2%; Late: 94.7%). Of the early phase studies, 23.1% were

identified as using adaptive methods and 41.1% of late phase studies used adaptive methods. Of the drug treatment studies, protein kinase inhibitors were primarily used in both phases of studies (Early: 17.4%; Late: 17.5%). Early phase studies tended to be single arm studies (54.9%) where late phase studies tended to have two treatment arms (79.0%). Primary funding for early phase studies came from industry (50.0%) and from other sources (41.3%), whereas late phase studies got most of their primary funding from industry (88.4%). Most of the studies were initiated from 2011 through 2014, but this was also likely related to the randomization sequencing where a greater weight of selection was placed on more recently initiated studies to increase the likelihood of the inclusion of adaptive studies as noted by other researchers (Hatfield et al., 2016). A description of study initiation by design method (early versus late) is shown in Figure 4. Figure 5 includes the study design by year of initiation. Most studies were completed (57.5%) or active and not recruiting (24.9%) for both study phases (early versus late).

For studies included in the analysis, if an early or late study was stopped prematurely, the primary reason was due to futility or safety (Early: 52.8%; Late: 40.0%). As expected, late phase studies had greater duration in comparison to early phase studies, though the median difference was only 2 months (28.0 versus 30.0 months respectively). The time until results were entered in the Clinical Trials database was comparable as well (Early: 13.0 months; Late: 12.0 months), though the Clinical Trials database result reporting guideline is likely the driving factor for the similarities of time until results availability. Results were favorable for 59.1% of early phase studies and 52.6% of late

stage studies, of which most clinical study results included were not reported to be published regardless of study stage (Early: 90.9%, Late: 89.5%, Overall: 90.6%).

As expected, late stage studies median study size were substantially larger than early phase studies (Median sample size Early: 22; Late: 158). Masking, data monitoring committees and treatment randomization was more often used in late stage studies, in comparison to early studies. Most studies regardless of early or late phase had a surrogate clinical primary endpoint. Regardless of phase, most studies also included biomarker assessments as a study endpoint. Median duration between early and late phase studies was 27.2 months (range: 10.8 to 50.0). Additional study characteristics are shown in Table 12.

Table 12

Summary of Study Characteristics Included in Analysis

	Early phase[a]	Late phase[a]	Total
Number of studies, N (%)	286 (75.1)	95 (24.9)	381 (100.0)
Phase[a] (N, %)			
Pilot/NA	6 (2.1)	0	6 (0.2)
Early Phase I	3 (1.1)	0	3 (0.8)
I	28 (9.8)	0	28 (7.3)
II	199 (69.2)	0	198 (52.0)
III	0	88 (92.6)	88 (23.1)
IV	0	7 (7.4)	7 (1.8)
Combined phases	51 (17.8)	0	51 (13.4)
0/I	0	0	0
I/II	46 (16.1)	0	46 (12.1)
II/III	5 (1.8)	0	4 (1.1)
Not reported	0	0	0
Study type (N, %)	0	0	0
Expanded access			
Interventional	285 (99.7)	93 (97.9)	378 (99.2)
Observational	0	0	0
Observational – patient registry	0	0	0
Not reported	1 (0.4)	2 (2.1)	3 (0.8)
Treatment classification[b] (N, %)			
Drug	275 (96.2)	90 (94.7)	365 (95.8)
Device	6 (2.1)	1 (1.1)	7 (1.8)
Biological/vaccine	0	0	0
Procedure/surgery	24 (8.4)	0	24 (6.3)
Radiation	17 (5.9)	4 (4.2)	21 (5.5)
Other	30 (10.5)	4 (4.2)	34 (8.9)
Combination treatment	145 (50.7)	49 (51.6)	194 (50.9)
Drug treatment[cd] (N, %)			
Cisplatin	26 (4.0)	1 (0.6)	27 (3.3)
Paclitaxel	16 (2.5)	11 (6.4)	27 (3.3)
Cyclophosphamide	25 (3.9)	0	25 (3.1)
Trastuzumab	14 (2.2)	10 (5.8)	24 (2.9)
Docetaxel	14 (2.2)	8 (4.7)	22 (2.7)
Gemcitabine	18 (2.8)	4 (2.3)	22 (2.7)
Carboplatin	16 (2.5)	5 (2.9)	21 (2.6)
Fluorouracil	16 (2.5)	4 (2.3)	20 (2.5)
Dexamethasone	14 (2.2)	5 (2.9)	19 (2.3)
Prednisone	11 (1.7)	5 (2.9)	16 (2.0)
Bortezomib	11 (1.7)	3 (1.8)	14 (1.7)
Capecitabine	12 (1.9)	2 (1.2)	14 (1.7)
Oxaliplatin	14 (2.2)	0	14 (1.7)
Aldesleukin	13 (2.0)	0	13 (1.6)

(continued)

Table 12 Continued

	Early phase[a]	Late phase[a]	Total
Drug treatment continued			
Bevacizumab	8 (1.2)	5 (2.9)	13 (1.6)
Fludarabine	13 (2.0)	0	13 (1.6)
Irinotecan	11 (1.7)	2 (1.2)	13 (1.6)
Lenalidomide	10 (1.6)	3 (1.8)	13 (1.6)
Erlotinib	10 (1.6)	2 (1.2)	12 (1.5)
Leucovorin	9 (1.4)	2 (1.2)	11 (1.3)
Abiraterone	9 (1.4)	1 (0.6)	10 (1.2)
Pertuzumab	7 (1.1)	3 (1.8)	10 (1.2)
Other[d]	338 (52.5)	93 (83.0)	431 (52.9)
Drug classification[cd] (N, %)			
Protein kinase inhibitors	112 (17.4)	30 (17.5)	142 (17.4)
Monoclonal antibodies	77 (12.0)	29 (17.0)	106 (13.0)
Taxanes	43 (6.7)	21 (12.3)	64 (7.9)
Platinum compounds	57 (8.9)	6 (3.5)	63 (7.7)
Pyrimidine analogues	50 (7.8)	10 (5.8)	60 (7.4)
Other antineoplastic agents	40 (6.2)	13 (7.6)	53 (6.5)
Nitrogen mustard analogues	32 (5.0)	1 (0.6)	33 (4.0)
Other	24 (3.7)	2 (1.2)	26 (3.2)
Corticosteroids	16 (2.5)	6 (3.5)	22 (2.7)
Anti-androgens	13 (2.0)	7 (4.1)	20 (2.5)
Other immunosuppressants	14 (2.2)	4 (2.3)	18 (2.2)
Folic acid metabolite	13 (2.0)	3 (1.8)	16 (2.0)
Glucocorticoids	11 (1.7)	5 (2.9)	16 (2.0)
Drug classification continued			
Interleukins	14 (2.2)	0	14 (1.7)
Purine analogues	13 (2.0)	0	13 (1.6)
Other hormone antagonists and related agents	9 (1.4)	1 (0.6)	10 (1.2)
Other[d]	106 (16.5)	33 (19.3)	139 (17.1)
Number of treatment arms (N, %)			
1	157 (54.9)	11 (11.6)	168 (44.1)
2	80 (28.0)	75 (79.0)	155 (40.7)
3	27 (9.4)	4 (4.2)	31 (8.1)
4	9 (3.2)	2 (2.1)	11 (2.9)
≥ 5	12 (4.2)	1 (1.1)	13 (3.4)
Not reported	1 (0.4)	2 (2.1)	3 (0.8)
Primary funding source			
Industry	143 (50.0)	84 (88.4)	227 (59.6)
Public	23 (8.0)	1 (1.1)	24 (6.3)
Both	0	0	0
Other	118 (41.3)	8 (8.4)	126 (33.1)
Not reported	2 (0.7)	2 (2.1)	4 (1.1)

(continued)

Table 12 Continued

	Early phase[a]	Late phase[a]	Total
Additional funding source			
Industry	42 (14.7)	1 (1.1)	43 (11.3)
Public	16 (5.6)	3 (3.2)	19 (5.0)
Other	2 (0.7)	4 (4.2)	6 (1.6)
More than one	5 (1.8)	1 (1.1)	6 (1.6)
Not reported/none	221 (77.3)	86 (90.5)	307 (80.6)
Year study registered			
1999	1 (0.4)	2 (2.1)	3 (0.8)
2005	2 (0.7)	2 (2.1)	4 (1.1)
2006	2 (0.7)	4 (4.2)	6 (1.6)
2007	8 (2.8)	6 (6.3)	14 (3.7)
2008	6 (2.1)	1 (1.1)	7 (1.8)
2009	8 (2.8)	0	8 (2.1)
2010	9 (3.2)	0	9 (2.4)
2011	32 (11.2)	8 (8.4)	40 (10.5)
2012	94 (32.9)	25 (26.3)	119 (31.2)
2013	68 (23.8)	22 (23.2)	90 (23.6)
2014	35 (12.2)	21 (22.1)	56 (14.7)
2015	14 (4.9)	2 (2.1)	16 (4.2)
2016	7 (2.5)	2 (2.1)	9 (2.4)
Study status (n, %)			
Active, not recruiting	54 (18.8)	41 (43.2)	95 (24.9)
Completed	177 (61.9)	42 (44.2)	219 (57.5)
Suspended	0	0	0
Unknown	2 (0.7)	2 (2.1)	4 (1.0)
Terminated	53 (18.5)	10 (10.5)	63 (16.5)
Reason stopped			
Accrual problem	6 (11.3)	2 (20.0)	8 (2.1)
Business decision	7 (13.2)	1 (10.0)	8 (2.1)
Complete	1 (1.9)	0	1 (0.3)
Funding problem	1 (1.9)	0	1 (0.3)
Futility or safety concerns	28 (52.8)	4 (40.0)	32 (8.4)
Other reason	3 (5.7)	2 (20.0)	5 (1.3)
Not reported	7 (13.2)	1 (10.0)	8 (2.1)
Duration of study (months)			
Median	28.0	30.0	28.0
(Min, max, STD, N)	(3, 185, 18.0, 282)	(9, 261, 35.5, 93)	(3, 261, 23.8, 375)
Time until results entered in database (months)			
Median	13.0	12.0	13.0
(Min, max, STD, N)	(1, 125, 20.8, 285)	(3, 87, 14.3, 94)	(1, 125, 19.5, 379)

(continued)

Table 12 Continued

	Early phase[a]	Late phase[a]	Total
Outcome of study			
Favorable	169 (59.1)	50 (52.6)	219 (57.5)
Equivalent	4 (1.4)	7 (7.4)	11 (2.9)
Unfavorable	113 (39.5)	38 (40.0)	151 (39.6)
Journal publication			
Yes	43 (15.0)	10 (10.5)	53 (13.9)
No	241 (84.3)	83 (87.4)	324 (85.0)
Unknown	2 (0.7)	2 (2.1)	4 (1.1)
Results published[e]			
Yes	24 (8.4)	8 (8.4)	32 (8.4)
No	260 (90.9)	85 (89.5)	345 (90.6)
Unknown	2 (0.7)	2 (2.1)	4 (1.1)
Design characteristics			
Actual – median (Min, max, STD, N)	22 (1, 343, 47.9, 284)	158 (8, 2091, 275.5, 94)	36.0 (1, 2091, 167.9, 378)
Study primary purpose			
Basic science	0	0	0
Diagnostic	0	0	0
Health services	0	0	0
Research			
Other	0	0	0
Prevention	0	0	0
Screening	0	0	0
Study primary purpose continued			
Supportive care	0	0	0
Treatment	285 (99.3)	93 (97.9)	378 (97.7)
Not reported	2 (0.7)	2 (2.1)	4 (1.0)
Study design			
Adaptive	66 (23.1)	39 (41.1)	106 (27.6)
Traditional	220 (76.9)	56 (59.0)	276 (72.4)
Masking			
No	258 (90.2)	57 (60.0)	315 (82.7)
Yes	26 (9.1)	36 (37.9)	62 (16.3)
Type of masking			
Single	3 (11.5)	1 (2.8)	4 (6.5)
Double	10 (38.5)	14 (38.9)	24 (38.7)
Triple	4 (15.4)	3 (8.3)	7 (11.3)
Quadruple	9 (34.6)	18 (50.0)	27 (43.6)
Not reported	2 (0.7)	2 (2.1)	4 (1.1)
Subject	25 (96.2)	34 (94.4)	59 (95.2)
Caregiver	13 (50.0)	21 (58.3)	34 (54.8)
Investigator	22 (84.6)	33 (91.7)	55 (88.7)
Outcome assessor	11 (42.3)	20 (55.6)	31 (50.0)

(continued)

Table 12 Continued

	Early phase[a]	Late phase[a]	Total
Randomization			
Nonrandomized	52 (18.2)	2 (2.1)	54 (14.2)
Randomized	80 (28.0)	80 (84.2)	160 (42.0)
Not reported	154 (53.9)	13 (13.7)	167 (43.8)
Intervention model			
Crossover	2 (0.70)	1 (1.1)	3 (0.8)
Factorial	1 (0.4)	0	1 (0.3)
Parallel	92 (32.2)	78 (82.1)	170 (44.6)
Sequential	3 (1.1)	1 (1.1)	4 (1.1)
Single group	184 (64.3)	11 (11.6)	195 (51.2)
Not reported	4 (1.4)	4 (4.2)	8 (2.1)
Data monitoring committee (DMC)			
Yes	126 (44.0)	54 (56.8)	180 (47.2)
Type of primary endpoint[e]			
Surrogate	271 (94.8)	73 (76.8)	344 (90.3)
Clinical	286 (99.7)	95 (100)	380 (99.7)
Clinical and surrogate	269 (94.1)	73 (76.8)	342 (89.8)
Biomarker endpoint included[e]			
No	1 (0.40)	4 (4.2)	5 (1.3)
Yes	285 (99.6)	91 (95.8)	376 (98.7)
Time between late-and early-phase studies (months)[f]			
Median (Q1, Q3)	-	-	27.2 (10.8, 50.0)
(Min, Max, STD, N)			(0.2, 111.6, 26.3, 251)

a. Study phase of less than III as well as seamless studies of phase II/III were considered early since they included an early phase component.
b. A single study could be classified in multiple treatment interventions, thus percentages will add up to greater than 100%.
c. Overall counts within a treatment classification of less than 10 were combined into the 'Other' category. A complete list of experimental treatments and drug classifications can be found in the Appendix C.
d. As study treatments could include combination therapy, a single study could be represented in multiple rows.
e. Results could be published in a journal, or presented at a conference etc.
f. A surrogate endpoint is defined to be a primary endpoint that is not overall survival. Clinical endpoint is defined to be the primary endpoint that measures the clinical benefit of the treatment such as response or progression free survival. A biomarker endpoint is defined to be a biological measure that predicts a clinical outcome (primary or secondary).
g. Only positive time between studies (early phase study is initiated before the results of late phase study) are summarized.

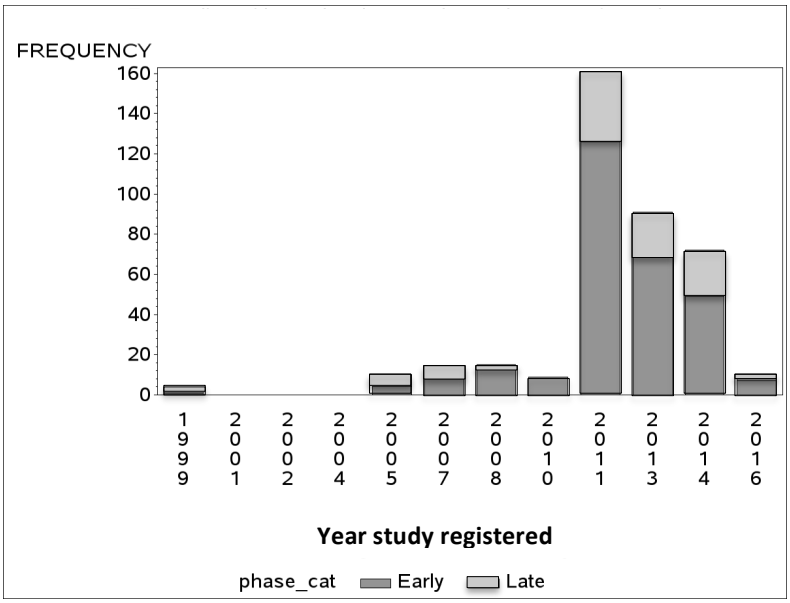

Figure 4. Study phase (early versus late) by year.

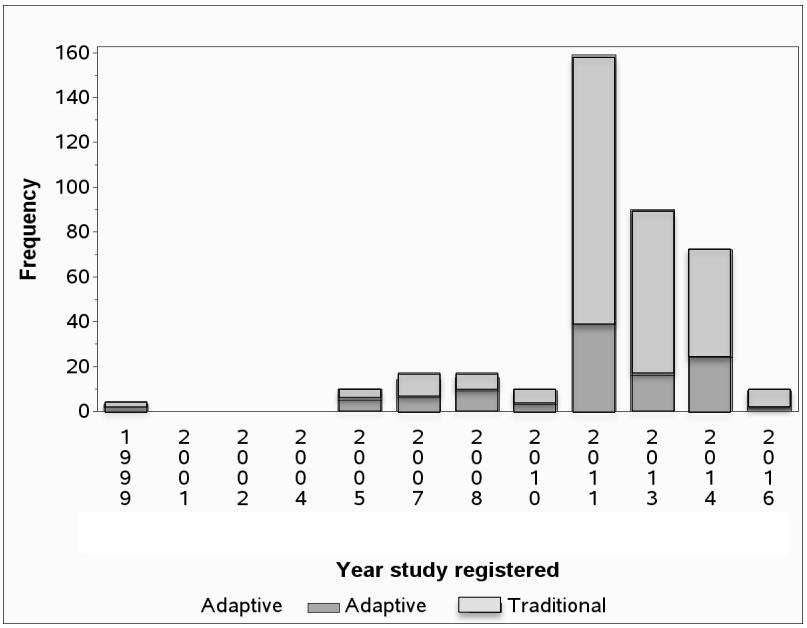

Figure 5. Study design (traditional versus adaptive) by year.

Changes in Data Analysis Plan

While the study outcome was collected in three categories (favorable, equivalent and nonfavorable), there were limited equivalence studies (N = 11) identified in the random analysis sample. In equivalency studies, a favorable result would be equivalence (as noted in Chapter 3, Research Design and Rationale Section). As a result, studies with outcomes that were deemed equivalent will be included as favorable in this analysis.

Due to lack of variability in the analysis dataset with respect to treatment classification, a modification was made to the third planned hypothesis of assessing the influence of treatment classification. The majority of studies (95%) were drug treatment studies and frequencies of device (1.8%), procedure/surgery (6.3%) or radiation (5.5%) studies reported were quite low. As a result, drug classification (include the proportions in each category) was assessed in the model rather than treatment classification.

The clinical trials database includes disease condition or conditions, with the Medical Subject Heading (MeSh) terms that were included within the same database field. However, programmatic identification of the disease condition was inconsistent for approximately 26% of the studies included in the analysis. This issue was identified at analysis, after the random sample of the Clinical Trials database was determined. As a result, the planned stratified random sample based on the disease conditions being correctly identified, were no longer balanced. Programmatic extraction was also difficult when there were multiple conditions within a single study due to the multiple MeSH terms. Though not in the original analysis plan, all programmatically identified disease conditions included in the random sample, I manually reviewed and verified.

In addition, a greater rate of studies that were reported to have results, were actually incomplete studies. The primary reason for these studies being incomplete was due to lack of accrual. Even with a random sample expanded to 537 rather than the planned 450, only 381 studies were included in this study with complete results. As shown in Table 13, a reduced sample size and varying percentage between the groups of interest (traditional versus adaptive), led to varying power than originally computed. Nonetheless, all randomly identified studies with available complete results were included in the analysis. See Figure 3, Tables 8 and 9 for power estimates as well as a summary of studies included and excluded from the analysis.

Table 13

Actual Sample Size with Assumptions for Proportion Positive Outcome, Power and Type I Error

		Proportion positive outcomes		Sample size	
Power (%)	Type 1	Traditional	Adaptive	Traditional	Adaptive
94	0.05	0.22	0.46	220	66
80	0.05	0.30	0.50	220	66
40	0.05	0.25	0.37	220	66
32	0.05	0.22	0.32	220	66
29	0.05	0.27	0.37	220	66
81	0.05	0.22	0.46	154	42
60	0.05	0.30	0.50	154	42
28	0.05	0.25	0.37	154	42
22	0.05	0.22	0.32	154	42
19	0.05	0.27	0.37	154	42
50	0.05	0.22	0.46	94	22
33	0.05	0.30	0.50	94	22
14	0.05	0.25	0.37	94	22
11	0.05	0.22	0.32	94	22
10	0.05	0.27	0.37	94	22
85	0.05	0.22	0.46	185	47
66	0.05	0.30	0.50	185	47
31	0.05	0.25	0.37	185	47
24	0.05	0.22	0.32	185	47
21	0.05	0.27	0.37	185	47

The initial data analysis plan was to include studies that were initiated January 1, 2000 to December 31, 2016. In the end, studies included, were any oncology intervention studies that had results available at extraction regardless of initiation date. The intervention studies included in this analysis were initiated after November 1, 1999 to December 31, 2016.

The random forest analysis of the time between the early and late stage study was not conducted, as the procedure within SAS Enterprise Miner was not available. As a

result, the quartiles of the duration of time were computed and included in the analysis model.

Univariate Analysis of Covariates

The univariate analyses included the following variables: experimental treatment, cancer type, experimental drug classification, duration between studies (months), study funding, type of endpoint, sample size and biomarker endpoint. Due to sparse or lack of variability in the data, robust estimates could not be computed for some independent variables included in the model. Of the univariate analyses, the only covariate, duration of time between studies was found to be overall statistically significant with late phase outcome as the dependent variable (p value=0.03). There were lower odds of having a favorable outcome for older studies (greater than Q3) in comparison to more recent studies (less than Q1) (OR: 0.33; 0.14, 0.77). This indicates that there were higher odds of less favorable outcomes for studies greater than Q3 in comparison to studies less than Q1 (OR: 0.33; 95% CI: 0.14, 0.77). Table 14 provides additional details of the univariate covariate analysis with late phase results outcome. Figure 6 also provides a visual of the odds ratio for the univariate analyses conducted.

Table 14

Univariate Covariate Analysis with Late Stage Results as Outcome

Univariate covariate	DF	Estimate	Standard error (SE)	Odds ratio (OR) and 95% confidence interval	Overall p value/ p value
Phase					
NA/Phase 1	1	-0.48	0.46	0.62 (0.25, 1.53)	0.30
Phase I / Phase II	1	0.43	0.34	1.54 (0.79, 3.01)	0.21
Phase II (ref)	1	---	---	---	---
Experimental treatment classification	---	NE	NE	NE	Most studies were drug trials, so limited variability for this analysis.
Adaptive	1	-0.17	0.18	0.71 (0.36, 1.42)	0.33
Cancer type					0.19
Breast	1	-0.31	0.36	0.43 (0.08, 2.38)	0.39
Colorectal	1	0.29	0.49	0.78 (0.12, 5.02)	0.55
Head and neck	1	0.14	0.67	0.67 (0.08, 5.54)	0.84
Lung	1	-0.56	0.35	0.33 (0.06, 1.84)	0.12
Multiple myeloma	1	1.12	0.61	1.79 (0.24, 13.4)	0.06
Neoplasms, solid	1	0.54	0.64	1.00 (0.13, 7.89)	0.39
Ovarian	1	0.04	0.40	0.61 (0.10, 3.53)	0.92
Pancreatic	1	-0.56	1.28	0.33 (0.01, 8.18)	0.67
Prostatic	1	-1.25	0.59	0.17 (0.02, 1.23)	**0.03**
Other (ref)	---	---	---	---	---
Experimental drug classification					
Anthracyclines and related substances	---	NE	NE	NE	0.97
Anti-androgens	---	NE	NE	NE	0.99
Corticosteroids	---	NE	NE	NE	0.97
Folic acid metabolite	---	NE	NE	NE	0.99
Glucocorticoids	---	NE	NE	NE	0.98
Monoclonal antibodies	---	NE	NE	NE	0.99
Nitrogen mustard analogues	---	NE	NE	NE	0.98
Other alkylating agents	---	NE	NE	NE	0.99
Other antineoplastic agents	---	NE	NE	NE	0.97
Other hormone antagonists and related agents	---	NE	NE	NE	0.98
Other immunosuppressants	---	NE	NE	NE	0.98
Platinum compounds	---	NE	NE	NE	0.99
Protein kinase inhibitors	---	NE	NE	NE	0.98
Pyrimidine analogues	---	NE	NE	NE	0.99
Taxanes	---	NE	NE	NE	0.99
Other (ref)	---	---	---	---	---

(continued)

Table 14 Continued

Univariate covariate	DF	Estimate	Standard error (SE)	Odds ratio (OR) and 95% confidence interval	Overall p value/ p value
Duration between studies (months)					**0.03**
Greater than Q3	-0.60	1	-0.60	0.25	**0.02**
Between Q2-Q3	0.34	1	0.34	0.27	0.21
Between Q1-Q2	-0.24	1	-0.24	0.25	0.34
Less than Q1 (ref)	---	---	---	---	---
Study funding					0.36
Industry	-0.28	1	-0.28	0.43	0.52
NIH	0.13	1	0.13	0.82	0.87
Other (ref)	---	---	---	---	---
Sensitivity variables					
Type of endpoint					
Surrogate	NE	---	NE	NE	NE
Clinical	NE	---	NE	NE	NE
Sample size					0.44
	---	1	-0.004	0.004	0.44
	0.004				
Biomarker					.056
Yes	0.21	1	0.21	0.35	0.56
No (ref)	---	---	---	---	---

a. Reference groups for the odds ratio and p value comparison are identified by (ref).
Favorable/unfavorable: 120/76

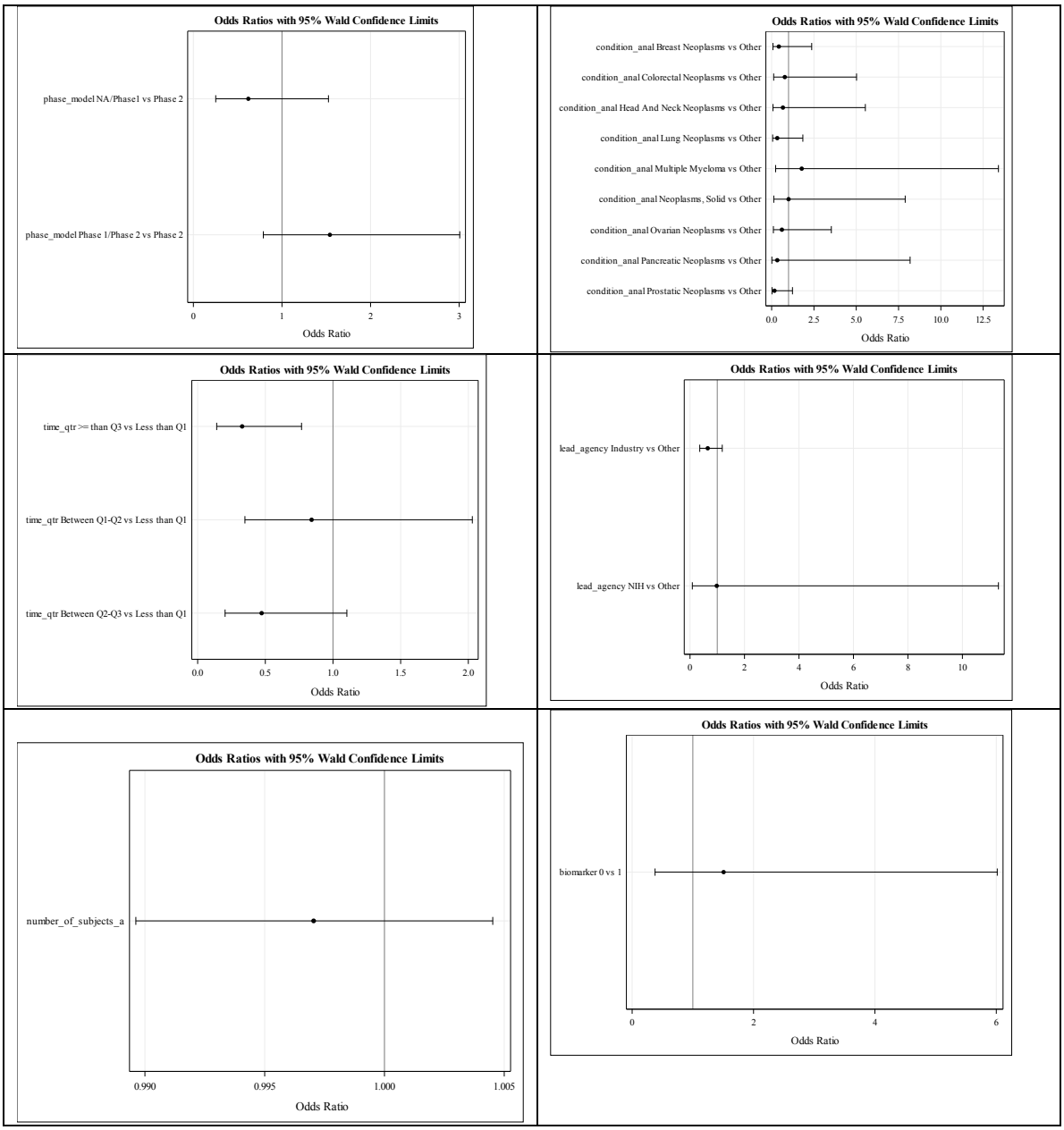

Figure 6. Odds ratio and corresponding 95% CI for univariate analyses.

Table 15 assesses the relationship between the early phase adaptive classification, the covariate and the late stage study outcome. Due to lack of variability some statistical estimates could not be computed specifically experimental treatment and drug

classification and type of endpoint. Estimates and significance of the covariates were similar in value and direction as what was seen in Table 14 (univariate analysis with late stage results as outcome).

Table 15

Influence of Specified Covariate on Adaptive Design Estimates in a Bivariate Model with Late Stage Results as Outcome

Univariate covariate	DF	Estimate	Standard error (SE)	Odds ratio[b] (OR) and 95% confidence interval (CI)	p value
Phase	1	-0.38	0.37	0.69 (0.33, 1.41)	0.31
Experimental treatment classification[a]	1	-0.31	0.36	0.73 (0.38, 1.50)	0.40
Cancer type	1	-0.35	0.40	0.71 (0.32, 1.54)	0.38
Experimental drug classification	1	-0.21	0.43	0.81 (0.35, 1.87)	0.62
Duration between studies (months)	1	-0.37	0.37	0.69 (0.34, 1.42)	0.31
Study funding	1	-0.49	0.37	0.62 (0.30, 1.27)	0.19
Sensitivity variables					
Sample size	1	-0.33	0.35	0.72 (0.36, 1.44)	0.35
Type of endpoint					
Surrogate[a]	1	-0.34	0.35	0.71 (0.36, 1.42)	0.33
Clinical[a]	1	-0.34	0.35	0.71 (0.36, 1.42)	0.33
Sample size	1	-0.33	0.35	0.72 (0.36, 1.44)	0.35
Biomarker	1	-0.32	0.35	0.73 (0.36, 1.46)	0.37

a. Not fully adjusted as the model was over-parameterized by the included covariate.
b. Odds ratio of adaptive design methods versus traditional, adjusted for specified covariate.
Favorable/unfavorable: 120/76

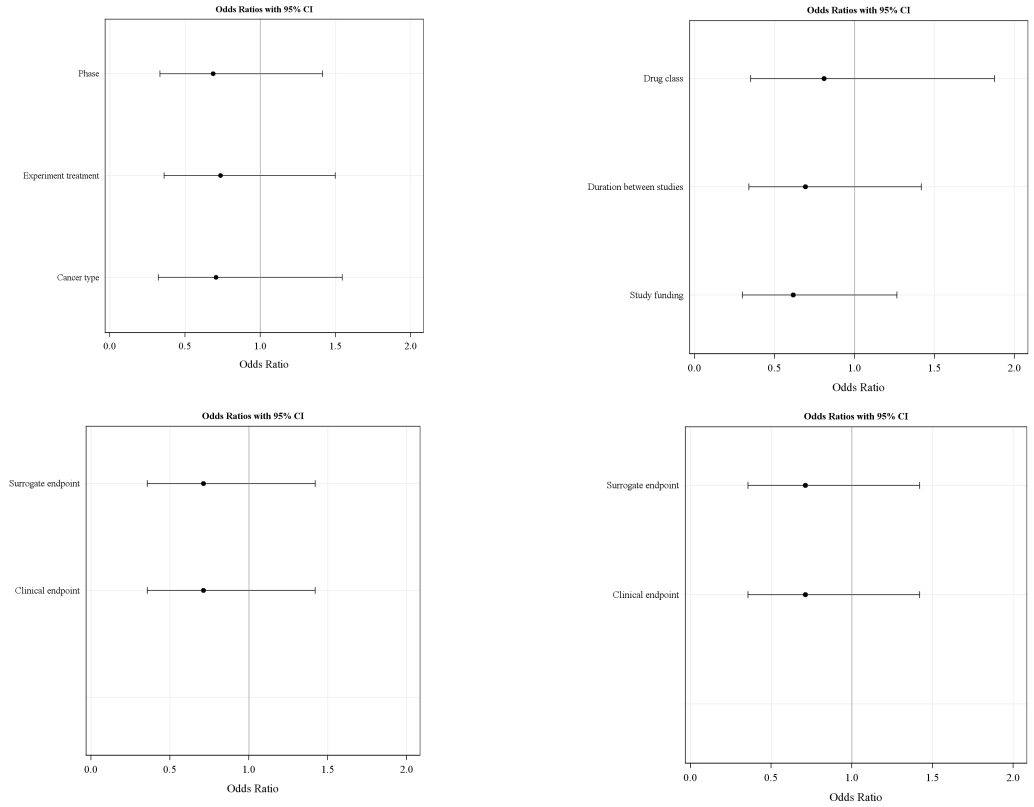

Figure 7. Odds ratio and corresponding 95% CI for univariate analyses adjusting for covariate adaptive design.

Bayesian Analysis

For all hypotheses driven logistic regression analyses conducted, a noninformative prior was utilized with a Gamerman sampling algorithm. Noninformative priors were selected as no similar studies had been conducted, thus no prior knowledge of the distribution or estimates were available. The Gamerman algorithm is known for its good performance in generalized linear models (Stokes, Chen and Gunes, 2014). A burn-in size of 2000 was used to reduce the chances of bias with initial estimates that may not be representative of the posterior distribution and Markov chain samples of 20,000 was used, unless otherwise specified. Thinning Markov chains to reduce the autocorrelation,

which may lead to biased estimates, was generally not used if the chain converged unless deemed necessary. Convergence assessment through multiple statistical tests were utilized including the effective sample sizes (ESS), Gelman-Rubin statistic, Geweke, Heidelberger-Welch, Raftery-Lewis statistics were used. Trace plots were used to visually assess good sample mixing and whether autocorrelation was acceptable for all parameters.

Results

Association between Adaptive Design in Early Phase sStudies and Outcome of Late Phase Studies (Hypothesis 1)

Early and late phase studies were merged based on the treatment of which 196 unique matched pairs (unique early and late study combinations based on treatment) were used in the analysis. Cancer type was not matched as researchers commonly share and use results of previously approved studies with common treatment to make decisions on the possibility of new indications (Institute of Medicine (US) Forum on Drug Discovery, 2010). Redundant combination studies with multiple populations were filtered for the analysis. Confounding variables in the initial model included study design, phase of study, duration between studies, type of endpoint and, primary funding source. The association between early phase design methods and late phase outcomes were assessed using Bayesian logistic regression methods with a noninformative prior.

The initial model was over-parameterized likely due to the lack of variability in some of the variables. The variables sample size, surrogate endpoint and clinical endpoint were dropped from the model due to over-parameterization and lack of variability. Study

funding was reduced from three levels to two as the category, NIH only had a frequency of 3. As evidenced by the initial trace plots and the evidence of possible nonconvergence, the burn-in was increased from 2000 to 3000 while the number of simulations remained the same (N = 20,000). The final trace plots looks satisfactory as well as the convergence test.

The Gelman-Rubin and Geweke convergence tests were re-assessed with 3000 burn-in simulations and moderate test statistics for the intercept and the adaptive parameter indicating that convergence of the Markov chain has been satisfied. The Raftery-Lewis test was also conducted and both the dependence factors for the intercept as well as the adaptive design variable and other confounding variables were the appropriate level, indicating that there is no concern with respect to correlation between samples. The Heidelberger-Welch stationarity and half-width diagnostics were conducted and all outcomes were satisfactory, indicating that correlation of samples was not a concern and the simulation sample size was appropriately large. The Effective Sample Size (ESS) was also assessed. The ESS was identified to be large enough for the intercept, adaptive design covariate and other confounding variables indicated good mixing. Additional details are shown in Table 16.

The trace plots diagnostics for the intercept, adaptive variable and the confounding variables were assessed. Convergence was further assessed via the trace plots and good mixing is apparent (Figure 8). The posterior autocorrelation was also assessed, by reviewing the trace plots and the drop off is rapid, indicating there is adequate mixing of the Markov chain. For additional details, see Figure 8. All diagnostics

of the model indicate that the model is stable and the posterior estimates can be assessed.

Details of all hypotheses model fit are in Table 16.

Table 16

Assessment Statistics for the First Hypothesis Model: Early Phase Adaptive Design Utilization and the Association with Late Phase Study Outcome (Burn-in=3000, MCMC simulations=20,000)

Model assessment test	Adaptive parameter	Other parameters[a]
Gelman-Rubin (estimate, 97.5% bound)	1.00	1.00
Geweke Z (*p*-value)	-0.11 (0.92)	-1.11 (0.27)
Heidelberger-Welch (stationarity test, half-width test)	Pass; Pass	Pass;Pass for all
Raftery-Lewis (dependence factor)	2.71	1.65
Effective sample sizes (range)	8405	(6907, 8186)

a. Confounding variables include: Duration between studies, primary funding source, type of endpoint (biomarker, surrogate, clinical) and phase of study.

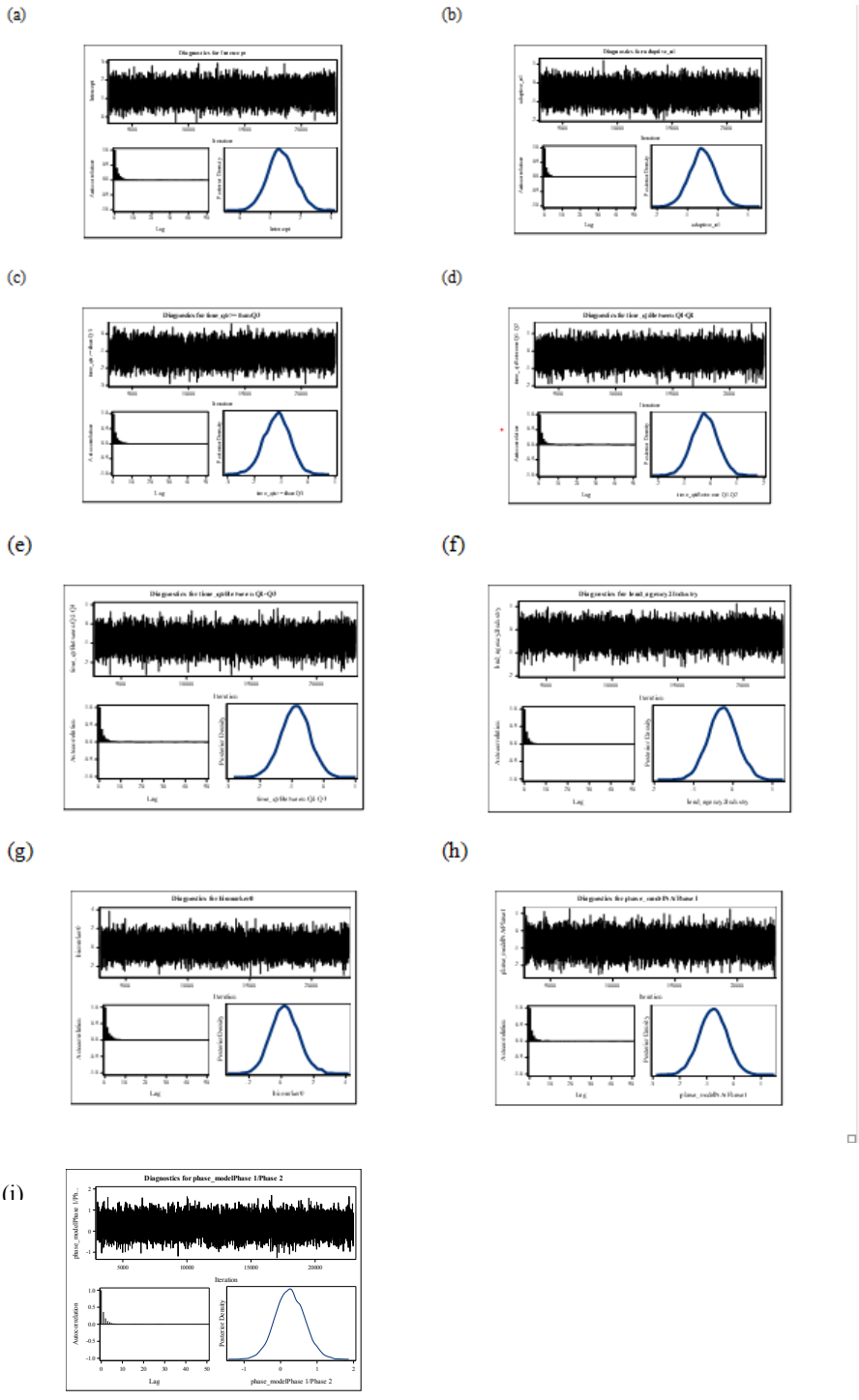

Figure 8. Trace plots for the Bayesian logistic regression analysis of the early phase covariate adaptive and the late phase study outcome for the intercept[a] and adaptive covariate[b].

The odds ratio estimate was computed using PROC PLM along with the 95% highest posterior density (HPD), the Bayesian equivalent of the frequentist confidence intervals. The odds ratio was 0.66 with a 95% HPD of 0.20 to 1.21 for adaptive design controlling for the other confounding variables in the model. In the adjusted model, while the estimate was not significant (95% HPD includes 1), there does appear to be a lower odds of a positive outcome in late stage studies when adaptive methods are utilized in early phase studies. Additional details can be found in Table 18.

Table 17

Fit Statistics Related to All Three Hypotheses

			Fit statistics	
Model	Seed	Description	DIC	pD
Hypothesis 1[a]	1234	Outcome (+/-) = Design Method (Adaptive/Traditional)	265.2	9.12
Hypothesis 2[b]	1234	Outcome(+/-) = Design Method (Adaptive/Traditional) + Condition + interaction	123.8	4.8
Hypothesis 3[c]	1234	Outcome(+/-) = Design Method (Adaptive/Traditional) + Drug Class + interaction	292.0	6.9

a. Adjusting for confounding variables: phase of study, biomarker endpoint, duration between studies and primary funding source.
b. Adjusting for confounding variables: disease condition and interaction term (adaptive * disease condition)
c. Adjusting for confounding variables: drug classification and interaction term (adaptive * drug classification)

Table 18

Posterior Estimate Summaries for Hypothesis 1: Early Phase Adaptive Design Utilization and the Association with Late Phase Study Outcome (Burn-in=3000, MCMC simulations=20,000)

Parameter	Mean	Standard deviation	HPD	Interval	OR (95% HPD)
Intercept	1.36	0.44	0.5020	2.2321	---
Adaptive	-0.51	0.42	-1.3120	0.3162	0.66 (0.20, 1.21)
Duration between studies					
Between Q1-Q2	-0.23	0.48	-1.1709	0.7018	0.89 (0.21, 1.78)
Between Q2-Q3	-0.89	0.45	-1.7513	0.0326	0.45 (0.12, 0.88)
Greater than or equal to Q3	-1.17	0.48	-2.1171	-0.2479	0.35 (0.08, 0.68)
Funding source - Industry	-0.27	0.37	-0.9431	0.4932	0.82 (0.30, 1.45)
Biomarker endpoint	0.22	0.86	-1.5080	1.8718	1.82 (0.09, 5.29)
Study phase					
NA/Phase1	-0.77	0.51	-1.7486	0.2357	0.53 (0.13, 1.09)
Phase 1/Phase 2	0.25	0.39	-0.5353	1.0158	1.39 (0.46, 2.54)

Favorable/unfavorable: 120/76

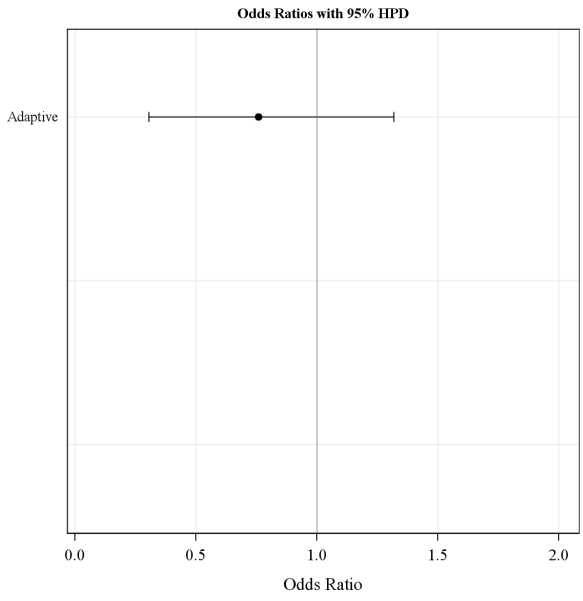

Figure 9. Odds ratio of favorable versus unfavorable results, when adaptive methods versus traditional design are adopted and corresponding 95% HPD controlling for confounding variables.

Association of Adaptive Design and Study Outcomes across Cancer Types (Hypothesis 2)

The second hypothesis of this study was to assess the association between early phase oncology design methods and the outcome of late stage clinical trials adjusting for specific cancer types and the interaction between adaptive design and cancer type. For this analysis, 116 unique early phase and late phase study combinations were identified that were matched on treatment and cancer type. Due to model over-parameterization and low frequencies for some cancer types, cancer types were limited to breast, lung and all other cancer types were combined in the 'other' category based on frequencies. Confounding variables in the initial model included study design, phase of study, duration between, type of endpoint, primary funding source, whether the study was combination treatment or not and the time between late and early phase studies. Combination treatment was included in the model as different treatments could be represented by the same study. All confounding variables were removed in the final model due to over-parameterization and lack of convergence. The interaction terms between adaption and cancer types were included in the model, and it was noted that the estimated parameter was in the positive direction, unlike the adaptive parameter estimate, which was negative, though not significant. The convergence tests appeared to be acceptable, though the trace plots indicated that autocorrelation remained a concern at lags greater than 10. The number of Monte Carlo simulations was increased to 35,000 and the burn in time was also increased to 5,000, which lead to more stable convergence

indicators. The final model included the variables adaptive, disease condition and the interaction term between the variables. Details can be found in Table 19.

Table 19

Assessment Statistics of the Model for the Second Hypothesis Model: Early Phase Adaptive Design Utilization and the Association with Late Phase Study Outcome Controlling for Disease Condition

Model assessment test	Adaptive parameter	Cancer type	Interaction terms[a]
Gelman-Rubin (estimate, 97.5% bound)	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)
Geweke Z (*p*-value)	-0.42 (0.68)	< 1.42 (>0.16)	>0.41 (>0.67)
Heidelberger-Welch (stationarity test, half-width test)	Passed/Passed	Passed/failed[b]	Passed for all
Raftery-Lewis (dependence factor)	5.42	<16.80	<12.83
Effective sample size	1210	<2478	<1211

a. All individual parameters have assessment statistics and can be found in Appendix D. The worst represented parameter statistic is displayed in the table.
b. Disease condition="Other" did not pass the Heidelberger-Welch half-width test.

The cancer type (lung versus breast cancer) in the model was significant in the model (β=-2.21, 95% HPD interval: -3.56, 0.89), indicating that the variable contributes to the association between early phase design methods and late stage outcome. Odds ratios were not computed for each individual term in the model due to the interaction term. Lung cancer and other cancer studies have a higher odds of positive outcome when adaptive designs are used in comparison to traditional methods, though not significant (Lung - OR: 3.29, 95% HPD: 0.13, 9.97; Other - OR: 7.65, 95% HPD: 0.15, 27.90). Breast cancer also had reduced odds of a positive outcome when adaptive methods are

used in comparison to traditional methods, though the results were once again not significant and with very wide HPD (OR: 0.04, 95% HPD: 0.0, >50). Additional details can be found in Table 20.

Table 20

Model Estimates for Hypothesis 2: Early Phase Adaptive Design Utilization and the Association with Late Phase Study Outcome Controlling for Disease Condition

Parameter	Mean	Standard deviation	HPD intervals	Odds ratio (OR)	OR HPD
Intercept	1.47	0.43	0.63, 2.27	---	---
Adaptive	-449	352	-1143, 0.6146	---	---
Cancer type[a]					
Lung	-2.21	0.70	-3.56, -0.89	---	---
Other	0.05	0.58	-1.07, 1.24	---	---
Interaction					
Adaptive (yes) x Lung	450	351.7	-1.22, 1142	---	---
Adaptive (yes) x Other	451	351.7	0.03, 1144	---	---
Adjusted estimates for adaptive versus traditional by cancer type					
Lung cancer[b]	0.74	0.95	-1.21, 2.50	3.29	(0.13, 9.97)
Other cancer[b]	1.36	1.15	-0.79, 3.60	7.65	(0.15, 27.90)
Breast cancer[b]	-499.4	351.7	-1143, 0.61	0.04	(0.0, >50)

Favorable/unfavorable: 85/31;

a. Reference groups: Combination treatment – No; Cancer type – Breast cancer;
b. Reference group: traditional design;

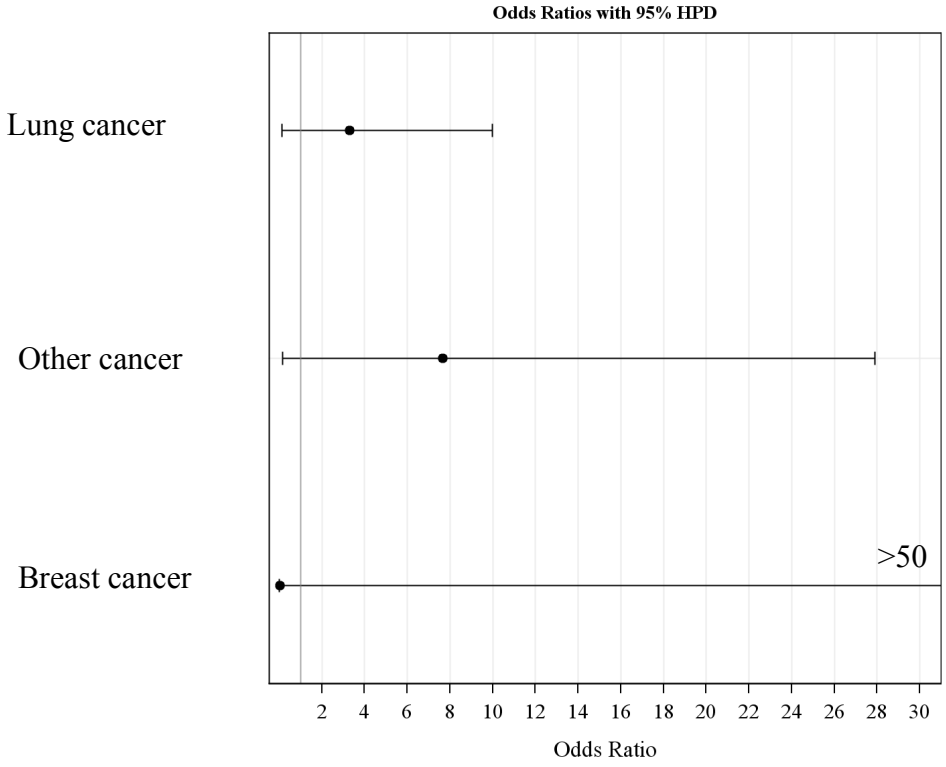

Figure 10. Odds ratio of favorable versus unfavorable results when adaptive versus traditional methods are adopted and corresponding 95% HPD.

Drug Treatment Classification (Hypothesis 3)

As previously noted, the random dataset primarily included drug treatments (95.8%). Unfortunately very few studies in early phase and late phase reported device use (Early: 2.1%; Late: 1.1%), procedures or surgery (Early: 8.4%; Late: 0.0%) or radiation therapy (Early: 5.9%; Late 4.2%). Due to these low frequencies a model could not be conducted to assess the relationship between treatment classifications, early phase treatment design and late phase outcome.

The analysis by treatment classification was limited as most of the studies included in this analysis were drug treatment. As such, the analysis of drug classification was conducted. For this analysis, 232 unique early phase studies combined with late phase studies were identified that were matched on treatment and drug classification. Drug classifications that had a frequency less than 20 were combined to avoid model over-parameterization. The final model included the drug classifications protein kinase inhibitors, monoclonal antibodies, taxanes and other to allow for parameter estimates. Confounding variables in the initial model included combination treatment (yes/no), duration between studies, primary funding agency, number of subjects included in the actual study, study endpoint (biomarker, surrogate, clinical) and study phase. Due to the lack of variability in the model and over-parameterization, the confounding variables (clinical (100%) and surrogate endpoint (100%) were also excluded from the model. Over-parameterization continued after multiple steps of modeling. As a result,

combination treatment, time between studies, primary funding, number of patients enrolled, biomarker and phase of the study were dropped from the model in order to assess the potential interaction of drug treatment classification and the relationship with study design method. Convergence was attained, however, autocorrelation between samples was elevated, indicating that mixing may be a concern (Minimum Geweke parameter *p*-value: 0.038) thus the burn in period was increased to 8,000 and the model was rerun. The trace plots of this model were assessed and the autocorrelation between samples did appear to drop off appropriately indicating convergence was attained. Thinning was also increased to 2 to reduce autocorrelation. The sampling method was also assessed to determine if Gibbs or Metropolis sampling algorithm would improve convergence and autocorrelation, but none performed as well as the Gamerman algorithm. The intercept was also removed from the model, as stationarity was not attained as indicated by the Heidelberger-Welch Half-width diagnostic test. In the final model, the monoclonal-adaptive and taxane-adaptive interaction terms, did not pass the Heidelberger-Welch half-width diagnostic test, however, all other convergence tests were satisfactory. As noted by researchers, Heidelberger-Welch can be significant, even when the trace plots appear acceptable. However, minor departures from stationarity should not be a concern, particularly when multiple convergence tests are being used to assess the quality of the model (Modlin, 2018). Table 21 provides additional details on the convergence statistics for this model.

In the final model, the covariate protein kinase was determined to be significant. Odds ratios for each individual covariate were not computed as a result of the interaction

term. Additional model details can be found in Table 22. The odds ratio for monoclonal antibody(MA), taxanes (TX) and other (OT) were positive, of which taxanes and other were significant (MA OR: 1.15, 95% HPD: 0.55, 1.79; TX OR: 2.75, 95% HPD: 1.01, 5.16; OT, OR: 3.23, 95% HPD: 1.58, 5.46). Protein kinase inhibitors (PK) had lower odds of a positive study when adaptive methods were used, though 95% HPD were very wide, thus indicating questionable robustness of the estimates (PK OR: 0.0005, 95% HPD: 0, >50). Details are included in Table 22 and Figure 11.

Table 21

Assessment of the Model for the Third Hypothesis Model: Early Phase Adaptive Design Utilization and the Association with Late Phase Study Outcome Controlling for Drug Classification

Model assessment test	Adaptive parameter	Drug classification	All other parameters[a]
Gelman-Rubin (estimate, 97.5% bound)	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)
Geweke Z (*p*-value)	>-1.63(>0.10)	<1.22 (> 0.36)	>-0.51 (>0.61)
Heidelberger-Welch (stationarity test, half-width test)	Pass/Pass	Pass/Pass	Pass/Passed for most[b]
Raftery-Lewis (dependence factor)	<3.07	<25.8	<7.75
Effective sample size	<4267	<4478	<4958

a. All parameters have assessment statistics and can be found in Appendix E. The most significant parameter statistic is displayed in the table.
b. The Heidelberger-Welch diagnostic stationarity test passed for all variables. The half-width test did not pass for the monoclonal-adaptive and taxane-adaptive interactions.

Table 22

Model Estimates for Hypothesis 3: Early Phase Adaptive Design Utilization and the Association with Late Phase Study Outcome Controlling for Drug Classification

Parameter	Mean	Standard deviation	HPD interval	Odds ratio (OR)	OR 95% HPD
Adaptive (yes)	1.15	0.58	0.10, 2.32	---	---
Drug classification[a]					
Monoclonal antibody	-0.90	0.41	-1.67, 0.09	---	---
Protein kinase inhibitor	-2.65	0.85	-4.43, -1.11	---	---
Taxanes	-0.18	0.39	-0.98, 0.58	---	---
Interaction					
Adaptive (yes) x Monoclonal antibody	-0.26	0.85	-1.90, 1.33	---	---
Adaptive (yes) x Protein kinase inhibitor	-779	601	-1952, 1.88	---	---
Adaptive (yes) x Taxanes	-0.03	1.00	-1.93, 2.04	---	---
Adjusted adaptive versus traditional estimates by drug classification[b]					
Monoclonal antibody	0.09	0.29	-0.47, 0.65	1.15	0.55, 1.79
Protein kinase inhibitor	-391.2	300.7	-977.1, -0.66	0.0005	0, >50
Taxanes	0.93	0.40	0.16, 1.72	2.75	1.01, 5.16
Other	1.12	0.32	0.50, 1.73	3.23	1.58, 5.46

Posterior summaries (simulations = 40,000)

Favorable/unfavorable: 150/82

a. Reference groups: Drug classification: Other;

b. Reference group: traditional design.

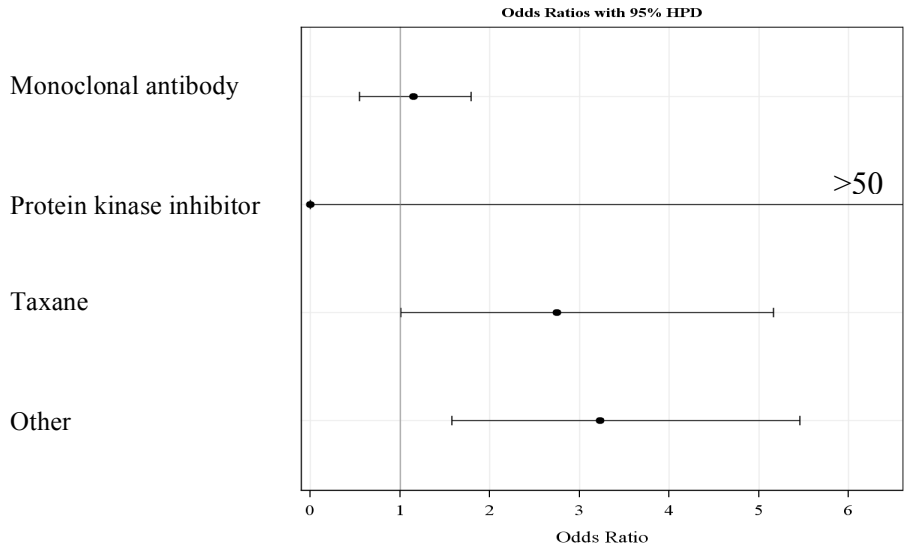

Figure 11. Odds ratio and corresponding 95% HPD for multivariate analyses including drug classification.

Sensitivity and Additional Analyses

Due to lack of variability in the data, full model analyses were limited and leading to over-parameterization. As such, a full model was conducted to assess the three hypotheses already assessed. Once again, studies that only had a positive difference in time so that the early phase studies occurred prior to the late stage studies were included in the analysis. Studies were merged and filtered by cancer type and treatment and 116 matched studies were included in the analysis. Clinical, surrogate and biomarker variables were removed from the model, as estimates for these variables could not be computed due to lack of variability thus model estimates were not computed.

In addition, a forward stepwise model was conducted to confirm any assessments that were made for the three hypotheses of this study. Akaike Information Criteria correct for bias (AICC) was used to determine which variable should be removed from the model. Finally, a forward stepwise model was conducted that included key study variables specifically adaptive design of the early phase study, drug class and cancer type. Once again, the AICC criterion was used to determine which variables should be included in the model.

The full analysis model was:

Phase III study results = early phase design (adaptive/traditional) + early phase + type of cancer + drug classification + duration between early and late phase studies +

(months) + study funding + actual sample size + biomarker endpoint in study + surrogate primary endpoint + combination treatment

Forward Stepwise Analysis

All Covariates Could be Excluded from the Model

A forward stepwise model was conducted that added one variable at a time automatically through SAS programming procedures (PROC HPGENSELECT). Variables were included based on the Schwartz Bayesian criterion (SBC) is used for the selection criteria which is a bias-corrected version of the Akaike's information criteria (AICC) (Schwarz, 1978). Cancer type and biomarker were entered into the model based on the AICC fit statistic, however, only cancer type was retained in the final model. Table 23 provides the final model and estimates.

As was found in the assessment of hypothesis 2 (Table 20), lung cancer was significant in the model (p-value=0.003) with an odds ratio of 0.18 (95% CI: 0.06, 0.56), which is in the same direction and a similar magnitude that was previously identified in the model that included adaptive methods. These estimates indicate that there is a reduced odd ratio of a positive study for lung cancer studies in comparison to breast cancer. There was also a reduced odds ratio of a positive study for studies that include 'other' cancers in comparison to breast cancer, though not significant. These results are also consistent with the results were identified in hypothesis 2 when study design was included in the model (Table 20). A summary of the analysis can be found in Table 23.

A similar forward stepwise analysis was conducted where key variables including adaptive design, disease condition and drug classification were forced to remain in the

final model. Similar results to all hypotheses and sensitivity analyses were noted and reported in Table 24.

Table 23

Forward Stepwise Model - Parameter Estimates

Parameter	DF	Estimate	Standard error	95% confidence limits (CL)		Odds ratio (OR)	95% OR CL		Chi-Square	Pr > ChiSq
Intercept	1	1.29	0.40	0.51	2.07	--	--	--	10.40	0.0013
Condition – lung neoplasms	1	-1.73	0.58	-2.88	-0.58	0.18	0.06	0.56	8.75	0.0031
Condition – other	1	0.37	0.54	-0.69	1.42	1.45	0.50	4.14	0.46	0.50

Table 24

Forward Stepwise Model Including Key Variables (adaptive, cancer type, drug classification) - Parameter Estimates

Parameter	DF	Esti-mate	Standard error	95% confidence limits (CL)		Odds ratio (OR)	95% OR CL		Chi-square	Pr > ChiSq
Intercept	1	2.27	0.91	0.50	4.04	--	--	--	6.29	0.01
Adaptive	1	0.66	0.70	-0.70	2.03	1.93	0.50	7.61	0.90	0.34
Drug classification										
Monoclonal antibodies	1	-1.13	0.90	-2.90	0.64	0.32	0.06	1.90	1.57	0.21
Platinum compounds	1	-0.61	1.07	-2.72	1.50	0.54	0.07	4.48	0.32	0.57
Pyrimidine analogues	1	0.91	1.12	-1.28	3.11	2.48	0.28	22.42	0.66	0.42
Taxanes	1	-0.69	0.83	-2.32	0.94	0.50	0.10	2.56	0.69	0.41
Cancer type										
Lung neoplasms	1	-2.47	0.89	-4.22	-0.72	0.08	0.01	0.49	7.63	0.01
Other	1	-0.82	0.94	-2.67	1.02	0.44	0.07	2.77	0.76	0.38

Table 25

Fit Statistics for Sensitivity Analysis Models

Model	AICC fit statistic
Forward stepwise model	125.0
Forward stepwise model including key variables (adaptive, cancer type, drug classification)	133.0

Summary

The clinical trial database was used to assess the association between design methods in early phase studies and the possible association with late phase outcomes in oncology studies. The association was also examined in categories of cancer type and drug classification. Five hundred thirty seven randomly selected studies where included in this analysis and were classified on their design methods as well as the final outcome of the study. Most early phase studies where phase II (69.2%) and most late phase studies were phase III (92.6%). The most common cancers included in early stage studies were lung (12.9%), colorectal (12.9%), breast neoplasms (12.4%) and multiple myeloma (12.6%). Multiple cancer types were included in 8.0 percent of early phase studies. Most studies were intervention (99.2%), drug studies (95.8%) and industry sponsored (59.6%). A variety of drug treatments were used of which protein kinase inhibitors were found to be the most frequent (17.4%). Of the early phase studies, 76.9% and 23.1% were classified as using traditional and adaptive design methods respectively. Most of the studies in early and late stage had a favorable or equivalent outcome (Early: 60.5%; Late: 60.0%).

Univariate analyses were conducted and adaptive design was not found to be associated with late stage outcome (p value=0.33). The only variable that was associated with late stage outcome was duration between studies. Studies that were greater than the Q3 apart in duration between studies had a reduced odds of a positive outcome in comparison to more recent studies (Q1 or less). No other variables were found to have an association with the late stage outcome, though specific cancer subtypes (multiple myeloma and prostatic) were found to be significant in comparison to other cancer types (*p*-values were less or equal to 0.05).

Bayesian logistic regression analyses methods were used to assess the association between early phase design methods (adaptive versus traditional) and the outcome of the late phase studies. Convergence statistics were assessed and no concerns were noted. There was no statistical association (OR: 0.66, 95% HPD: 0.20, 1.21) between early phase adaptive design studies and the outcome of late phase studies, adjusting for the potential confounding factors duration between studies, primary funding source, biomarker endpoint and study phase.

The association between early phase design studies and late phase outcome were assessed controlling for disease condition. In this model, while no cancer type was significant, lung cancer and other cancer had higher odds of positive outcome when adaptive methods were used in comparison to traditional (Lung - OR: 3.29, 95% HPD: 0.13, 9.97; Other - OR: 7.65, 95% HPD: 0.15, 27.90). Breast had reduced odds of a positive outcome when adaptive methods were used in comparison to traditional methods, though not significant (OR: 0.04, 95% HPD: 0.0, >50).

Similarly, a Bayesian logistic analysis was conducted assessing the association of early phase design and late stage outcome controlling for drug classification. Drug classifications taxanes and other drug classification were found to be marginally significant, with a higher odds of positive results when adaptive methods were adopted. The odds ratio for protein kinase inhibitors was close to 0 though the 95% HPD was quite wide, indicating questionable robustness of the estimate. Sensitivity and additional analyses were conducted and results were consistent with the primary analysis.

Chapter 5: Discussion, Conclusions and Recommendations

This quantitative observational study was used to investigate the potential association of early phase study design research outcomes that utilize traditional versus adaptive methods on late stage results in oncology clinical trials. As drug development and approval are based on cancer type and treatment classifications (Barabási et al., 2011; U. S. Food and Drug Administration, n.d.-b; Siddiqui & Rajkumar, 2012), these potential effect modifiers were examined to determine whether the association between use of adaptive versus traditional methods (i.e. 3+3) in early phase studies and the results outcome of the late stage study, was different in subgroups defined according to levels of these factors.

When controlling for cancer type as well as drug class including confounding variables, a nonstatistically significant difference in the odds of a positive study using adaptive design methods versus traditional was identified. Lung and other cancers appear to have a greater odds of positive results when adaptive methods are used versus traditional design methods. Breast cancer studies that used adaptive methods appeared to have reduced odds of a positive study when adaptive methods were used in comparison to traditional methods. Drug classifications, specification monoclonal antibodies, taxane and other appeared to have marginal impact increasing the odds of a positive study when adaptive methods were used in comparison to traditional methods, as odds ratios were greater than 1 for these drug classifications. While significant for taxanes and other, estimates should be interpreted with caution for both. Taxanes were estimated to be marginally significant as the lower 95% HPD is 1.01, just slightly higher than 1. In

addition, the other category was artificially created due to over-parameterization, thus likely heterogeneous.

This chapter includes a discussion and interpretation of the study results in conjunction with findings reported in the literature. Advantages and disadvantages related to limitations of this study are described. Recommendations for future research are also included.

Interpretation of the Findings

The purpose of the study was to investigate the potential associate of early phase study designs (adaptive versus traditional) and late stage results in oncology clinical trials. The association of cancer type and drug classification with design methods used for early phase oncology studies and the outcome of the late stage clinical trial was assessed using statistical modeling. A summary and interpretation of the findings can be found below.

Adaptive Design Methods

The association of adaptive design in early phase studies with the outcome of late phase studies adjusting for confounding variables including duration between studies, funding source, study endpoint and phase was assessed. There was no statistically significant association between adaptive design methods in early phase studies with the outcome of late phase studies. The odds ratio indicated that there was a reduced probability of positive late stage study when adaptive methods were used in comparison to traditional methods. While these results are not statistically significant, simulations indicated that adaptive methods were effective in treatment identification

(Ananthakrishnan et al., 2017). Further research in larger studies or targeted populations should be considered.

Drug Classification

While my initial intent was to assess treatment classifications in this study, the lack of heterogeneity in the data did not allow for this analysis. Drug classifications were assessed and within each classification, there was a marginal association between the study design method and study outcome. While this study marginally showed that early phase study design, late phase study outcome and drug classification were associated, oncology clinical trial development should be considered an iterative process of study development from early phase to late phase as indicated by the Framework for Design and Evaluation of Complex Interventions (DECI). This study did not definitively find an association between the variables of interest, perhaps a larger study or a study within a specific population may find a relationship between the study design and the study outcome.

Drug classification is just one component of the treatment development in oncology studies. Other confounding variables included in the initial model included type of endpoint, duration between studies, treatment classification and phase of study. Each identified and incorporated component indicates the complexity of oncology treatment development as well as the theoretical foundation of systems theory used to construct this study. While the change in drug treatment indicated marginal statistically significant difference when assessing the use of adaptive methods in early phase studies and the

influence on late stage results, the association of one component of clinical trial research with other components should not be ignored.

Cancer Type

By comparing adaptive versus traditional methods and the downstream clinical trial outcome, the possible benefits related to late stage study success were determined indicating that success in late stage studies may lead to improved treatment options for some cancers. Among breast cancer studies, there appeared to be lower odds of positive outcomes when adaptive methods were adopted in comparison to traditional methods. Lung and other cancers appear to have higher odds of positive outcomes when adaptive methods are used, though the results were not statistically significant. Using the study design and cancer type from early phase studies and assessing whether there is an association with the outcome of late phase studies is an indication of the complexities of oncology clinical trial development as well as the multiple considerations that must be addressed during treatment development. The complexity of the cancer type, phase of study and study endpoint should be considered to develop a quality study.

While the results of each hypothesis of this study were not statistically significant, further research should be conducted to assess the relationships. Further analyses of a larger sample size, a variety of cancer types and treatment plans would aid in identifying possible scenarios where adaptive methods are most effective. Using these results related to the first hypothesis of this study, in combination to the results of the second and third

hypothesis, additional research needs to be conducted to further understand the relationship of adaptive design method utilization and study outcome in oncology studies.

Limitations of the Study

All studies have limitations. However, the limitations may lead to future research, as more needs to be learned and addressed on the specific topic. The Clinical Trials database that I used in this research is quite extensive in the number of studies as well as the number of data points per study. However, limitations are evident in any archival data, specifically what was and what was not collected in the database as well as the format of the collected variables. For example, as has been reported by other researchers (Bothwell et al., 2018; Hatfield et al., 2016), the identification of adaptive studies in the Clinical Trial database is challenging and requires extensive manual review.

Clinical Trials Database

Many fields in the Clinical Trials database are free text format, thus varying text with the same message are frequently reported. The lack of consistent text made classification within variables difficult. As a result, in this research project, the classification of the study design methods, the study outcome, as well as basic study characteristics, such as phase of study and reason for termination were often supplemented with study publications or investigator contact for clarification when necessary. As such, while the database is informative on a study by study basis, aggregation and analysis of the database was very labor intensive as a result of formatting

and inconsistencies in text across studies. As a result of the key variables being labor intensive, a random sample of 381 studies were included in this analysis, rather than the 3040 oncology studies that were reported to have results. Increased power and the ability to assess the influence of early phase adaptive design utilization on late phase outcomes may be more feasible with a larger sample size.

As previous researchers have noted, overall the quality of the database is quite high (Hartung et al., 2014), however the capture of study design characteristics are inconsistently reported and limited with descriptive information (Anderson et al., 2015). Utilization of the clinical trials database as well as other databases collecting study details such as TrialTrove (Informa, n.d.) which was used in a recent publication by the Drug Information Association's Adaptive Design Scientific Working Group (ADSWG) in order to summarize adaptive design clinical trial utilization (Hartford et al., 2018). A larger sample size using systematic rather than random study identification may be incorporated in future research when multiple registration databases are used, which could simplify the labor-intensive process of study classification.

Delays in reporting of results in the Clinical Trials database have been noted (Anderson et al., 2015). In the current research, if a study did not have results, the study was automatically excluded from the analysis. Time until expected results were not assessed or summarized in this analysis for all studies. This exclusion criterion may have led to a biased sample.

Random Sample

This study included the analysis a random subset of archival ClinicalTrials.gov data. As the entire database was not used, however, when using a random sample, the underlying expectation is that the sample will be representative of the entire database and of clinical trials as a whole. However, the sample included a reduced percentage of adaptive studies in comparison to other similar analyses (Bothwell et al., 2018; Hatfield et al., 2016; Lin et al., 2016). Researchers have previously noted that between 37% to 42% of studies included in their analyses were adaptive (Bothwell et al., 2018; Hatfield et al., 2016). The random sample included in this analysis included a lower rate of adaptive studies (27.6%). However, unlike the other studies, this study includes phase I studies and strictly oncology research, which likely contributed to the lower rate of adaptive studies identified.

In addition, the random sample included a high percentage of positive outcome studies, which may indicate that bias may have been introduced when incomplete studies were removed from the analysis. Ideally, all oncology studies would be included in the analysis; however, due to the programmatic and manual data review, the sample size had to be limited for a timely analysis.

Model Development and Assumptions

One of the known benefits of adaptive studies is their ability to identify ineffective treatments earlier in comparison to traditional studies(Christopher S Coffey et al., 2012). To be included in the modeling of this analysis, early phase studies had to be matched on study characteristics, specifically treatment and/or disease condition

dependent on the hypothesis being assessed. As such, this study was conditional on treatments being available in late stage clinical trials, though not powered for this condition. This conditional sampling may be an area of future studies. The effectiveness of adaptive methods has been conducted in simulations (Bornkamp et al., 2007; Morgan et al., 2014) but a similar analysis has not been conducted on real data, as the feasibility is questionable.

A further limitation of this study is that the late stage outcome was strictly based on the primary endpoint. While the primary endpoint was negative, significant clinical secondary endpoints may have been statistically or clinically significant thus impacting future studies. Secondary endpoints were not a consideration of this study and may have blurred study impact on future development, regardless of the primary endpoint outcome. Once again, this may be an area where additional analyses should be conducted.

Recommendations

Considerations and Documentation of Design Methods

No study can address every issue related to a problem. Regardless of the study, further research is needed as a result of unanswered questions or questions raised during the study. With the increasing costs of healthcare and drug development, as well as the notable lag in the development of oncology clinical trials, adaptive design research should at a minimum be considered during the development phase of every clinical trial. While the design methods may not be appropriate for every situation, justification for not considering nontraditional designs should be considered and justification supporting or refuting the design should be well documented in trial records.

In addition, careful documentation and reporting of design methods (traditional versus adaptive) should be incorporated within the Clinical Trial database. With the definition of adaptive studies evolving over time, specifically when considering the recently released draft FDA adaptive design guidelines (U. S. Food and Drug Administration, 2018a) as well as the FDA master protocol guidelines for oncology studies (U. S. Food and Drug Administration, 2018b), documentation of the adaptive design definition that was utilized should also be reported in the database. With clearer documentation, further research could be conducted on the entire database and within diseases other than oncology. These additional analyses could aid in improving the understanding of the impact of adaptive methods as well as fully quantify their frequencies, benefits and limitations.

The possible inaccuracies introduced when self-reporting design methods need to be addressed. Additional collaboration with perhaps the NIH, AACT-CTTI and the subject matter experts such as the DIA Adaptive Design Scientific Working group (Drug Information Association, n.d.), should aid in appropriate and consistent classification of all studies within the database. In order to truly assess the impact of adaptive studies and the influence on treatment development, consistent classification of studies with input from the experts in the field is needed. In addition, classification within the database would reduce redundant work classifying studies by CT.gov users and researchers.

Study Endpoints

Further research is also needed to assess effective quality endpoints including composites in adaptive and tradition settings. While overall survival is considered the

gold standard in oncology, other endpoints may be more reflective of treatment efficacy for the population of interest. In this study, only the primary outcome was assessed and utilized in the analysis, regardless of what endpoint or type of endpoint was used in the study. However, study designs such as gatekeeper methods, mixed or composite endpoints or multiple primaries should also be considered when assessing the impact of adaptive methods on study outcomes. In addition, incorporating secondary endpoints in the analysis should be considered, as they may impact the future of the treatment, regardless of what the results of the primary endpoint is particularly in early phase studies.

Continued Education

A longitudinal qualitative study or a survey on clinical trial researchers should be conducted in the future to assess the researcher's perceived barriers on the adoption of adaptive methods. While this type of research has been conducted at a single point in time (Morgan et al., 2014), extended follow up would allow an assessment of perceived evolution of adaptive design barriers over time. Recently the Drug Information Association's Adaptive Design Scientific Working Group (ADSWG) published a survey about adaptive method use and changes (Hartford et al., 2018). The survey included questions about adaptive design use and perceived barriers.

In addition, as new researchers enter the field, this research could aid in illuminating similarities and differences within and across cohorts of interest, which would aid in future education development planning. The study could also lead to the development of appropriate adaptive design training developed for the specific audience

of interest, improving general knowledge on the topic at the university as well as continuing education programs. Using the research from the ADSWG on identified barriers would aid in training development.

Implications

Sharing of Results and the Data

In order to continue making progress in oncology clinical trials as well as reducing the incidence of similar treatment failures in targeted populations, quality meta-analyses, publicly available data of past studies need to be conducted on a regular basis and those results need to be shared. The Clinical Trials database as well as journal requirements for study registration has increased the opportunity for design, data and results sharing reducing the possible impact of publication bias. While individual clinical trials are important, appropriate meta-analyses need to be conducted in order to assess study trends with respect to safety, efficacy and noted trends and that have been reported. Additional information may be gleaned from the combined studies with the availability of increased information, sample size and power.

Sharing of information within programs is critical. However, sharing of information across programs, including with competitors is also important to continue moving the field forward and reduce the number of patients exposed to ineffective treatment. Researchers have reported results that furthered the treatment of oncology patients. The Clinical Trials results database should improve in sharing of results and information, however all studies should be required to share results to aid in future meta-analyses in all indications. These combined results should aid in moving the field forward

in oncology and allow for greater assessment of short term and long term efficacy and safety in the patient population or targeted treatment. Sharing of positive or negative results in a timely manner may reduce the rate of repeated study failures thus improve patient treatment options in the future.

Consideration of Adaptive Design Methods

While this analysis is only the beginning of assessing populations and treatments where adaptive methods can be used, further analyses are needed. In addition, researchers must continue to learn about new treatments, patient risks and predictive factors, but also newer design methods should be assessed. In order for treatment options to continue to be developed for the patient population, each component of the research field must be evaluated. The overall assessment and the adoption of possible new methods may lead to progress in every component of clinical research, including analysis and design methods. The quality of clinical research needs to continue to improve in order to increase the odds effective treatment options for the patient population.

Conclusion

This study assessed the possible relationship between early stage design methods and the outcome of late stage studies. The influence of cancer disease type and drug classification was also incorporated into the study and was found that there was a varying odds of a positive (lung and other cancer) and negative (breast) outcome in the late stage study when adaptive methods were utilized, though not significant or with wide credible intervals indicating estimates may not be robust. Drug classification appeared to have minimal influence on the relationship between early phase adaptive design utilization and

late stage study outcomes. These results indicate that further research in assessing design methods should be conducted. However, improved and consistent design reporting is needed to aid in the continued area of research.

Continuing to expand the Clinical Trials database as well as other databases by ex-United States regulatory bodies leads to shared information and additional analyses increasing knowledge. Meta-analyses and data sharing should increase patient and treatment related expertise in the field and hopefully improved treatment options for oncology patients as well as patients with other disease types as well. These types of analyses also utilized limited financial resources particularly in comparison to clinical research themselves and still continue to move the field forward with the aggregated analyses. However, registry database administrators should continue to assess the quality of their database as well as methods to improve the quality and accessibility to allow for expanded research.

References

Akaike, H. (2011). Akaike's information criterion. In M. Lovric (Ed.), *International encyclopedia of statistical science* (pp. 25-25). Berlin, Heidelberg: Springer.

Alavanja, M. C., Hoppin, J. A., & Kamel, F. (2004). Health effects of chronic pesticide exposure: cancer and neurotoxicity. *Annual Review Public Health, 25*, 155-197. doi:10.1146/annurev.publhealth.25.101802.123020

American Association for Cancer Research [AACR]. (n.d.). What is cancer? Retrieved from https://www.aacrfoundation.org/Pages/what-is-cancer.aspx

Ananthakrishnan, R., Green, S., Chang, M., Doros, G., Massaro, J., & LaValley, M. (2017). Systematic comparison of the statistical operating characteristics of various Phase I oncology designs. *Contemporary Clinical Trials Communications, 5*, 34-48. doi:10.1016/j.conctc.2016.11.006

Anderson, M. L., Chiswell, K., Peterson, E. D., Tasneem, A., Topping, J., & Califf, R. M. (2015). Compliance with results reporting at ClinicalTrials.gov. *New England Journal of Medicine, 372*(11), 1031-1039. doi:10.1056/NEJMsa1409364

Ardies, C. M. (2014). *Diet, exercise and chronic disease: The biological basis of prevention*. New York, NY: CRC Press.

Barabási, A.-L., Gulbahce, N., & Loscalzo, J. (2011). Network medicine: A network-based approach to human disease. *Nature Reviews Genetics, 12*(1), 56-68. doi:10.1038/nrg2918

Barker, A. D., Sigman, C. C., Kelloff, G. J., Hylton, N. M., Berry, D. A., & Esserman, L. J. (2009). I-SPY 2: an adaptive breast cancer trial design in the setting of

neoadjuvant chemotherapy. *Clinical Pharmacology & Therapeutics, 86*(1), 97-100. doi:10.1038/clpt.2009.68

Bauer, P., Bretz, F., Dragalin, V., König, F., & Wassmer, G. (2016). 25 years of confirmatory adaptive designs: Opportunities and pitfalls. *Statistics in medicine, 35*(3), 325-347. doi:10.1002/sim.6472

Bauer, P., & Einfalt, J. (2006). Application of adaptive designs–A review. *Biometrical journal, 48*(4), 493-506. doi:10.1002/bimj.200510204

Becvar, R. J., & Becvar, D. S. (2017). *Systems theory and family therapy: A primer* (3rd ed.). Lanham, MD: Rowman & Littlefield.

Berman, D. W., & Crump, K. S. (2008). A meta-analysis of asbestos-related cancer risk that addresses fiber size and mineral type. *Critical reviews in toxicology, 38*(sup1), 49-73. doi:10.1080/10408440802273156

Berry, D. A. (2011). Adaptive clinical trials: The promise and the caution. *Journal of Clinical Oncology, 29*(9), 606-609. doi:10.1200/JCO.2010.32.2685

Berry, D. A. (2012). Adaptive clinical trials in oncology. *Nature reviews Clinical oncology, 9*(4), 199-207. doi:10.1038/nrclinonc.2011.165

Biotechnology Innovation Organization [BIO]. (2016). Clinical development success rates 2006-2015. Retrieved from https://www.bio.org/sites/default/files/Clinical Development Success Rates 2006-2015 - BIO, Biomedtracker, Amplion 2016.pdf

Boissel, N., Auclerc, M.-F., Lhéritier, V., Perel, Y., Thomas, X., Leblanc, T., . . . Fegueux, N. (2003). Should adolescents with acute lymphoblastic leukemia be treated as old children or young adults? Comparison of the French FRALLE-93

and LALA-94 trials. *Journal of Clinical Oncology, 21*(5), 774-780. doi:10.1200/JCO.2003.02.053

Bornkamp, B., Bretz, F., Dmitrienko, A., Enas, G., Gaydos, B., Hsu, C.-H., . . . Neuenschwander, B. (2007). Innovative approaches for designing and analyzing adaptive dose-ranging trials. *J Biopharm Stat, 17*(6), 965-995. doi:10.1080/10543400701643848

Bothwell, L. E., Avorn, J., Khan, N. F., & Kesselheim, A. S. (2018). Adaptive design clinical trials: a review of the literature and ClinicalTrials.gov. *British Medical Journal Open, 8*(2). doi:10.1136/bmjopen-2017-018320

Boulding, K. E. (1956). General systems theory—The skeleton of science. *Management science, 2*(3), 197-208. doi:10.1287/mnsc.2.3.197

Brown, A. R., Gajewski, B. J., Aaronson, L. S., Mudaranthakam, D. P., Hunt, S. L., Berry, S. M., . . . Jawdat, O. (2016). A Bayesian comparative effectiveness trial in action: developing a platform for multisite study adaptive randomization. *Trials, 17*(1), 428. doi:10.1186/s13063-016-1544-5

Brown, C. H., Ten Have, T. R., Jo, B., Dagne, G., Wyman, P. A., Muthén, B., & Gibbons, R. D. (2009). Adaptive designs for randomized trials in public health. *Annual Review of Public Health, 30*, 1-25. doi:10.1146/annurev.publhealth.031308.100223

Buoen, C., Bjerrum, O. J., & Thomsen, M. S. (2005). How first-time-in-human studies are being performed: a survey of phase I dose-escalation trials in healthy

volunteers published between 1995 and 2004. *The Journal of Clinical Pharmacology, 45*(10), 1123-1136. doi:10.1177/0091270005279943

Burger, M., Catto, J. W., Dalbagni, G., Grossman, H. B., Herr, H., Karakiewicz, P., . . . Shariat, S. (2013). Epidemiology and risk factors of urothelial bladder cancer. *European urology, 63*(2), 234-241. doi:10.1016/j.eururo.2012.07.033

Butterfield, L. H., Disis, M. L., Khleif, S. N., Balwit, J. M., & Marincola, F. M. (2010). Immuno-oncology biomarkers 2010 and beyond: perspectives from the iSBTc/SITC biomarker task force. *Journal of translational medicine, 8*(1), 130. doi:10.1186/1479-5876-8-130

Campbell, M., Fitzpatrick, R., Haines, A., Kinmonth, A. L., Sandercock, P., Spiegelhalter, D., & Tyrer, P. (2000). Framework for design and evaluation of complex interventions to improve health. *British Medical Journal, 321*(7262), 694. doi:10.1136/bmj.321.7262.694

Carey, L., Winer, E., Viale, G., Cameron, D., & Gianni, L. (2010). Triple-negative breast cancer: disease entity or title of convenience? *Nature reviews Clinical oncology, 7*(12), 683-692. doi:10.1038/nrclinonc.2010.154

Centers for Disease Control and Prevention. (2017). Leading causes of death. Retrieved from http://www.cdc.gov/nchs/fastats/leading-causes-of-death.htm

Chapman, P. B., Liu, N. J., Zhou, Q., Iasonos, A., Hanley, S., Bosl, G. J., & Spriggs, D. R. (2017). Time to publication of oncology trials and why some trials are never published. *PLoS One, 12*(9), e0184025. doi:10.1371/journal.pone.0184025

Chen, A. Y., Jemal, A., & Ward, E. M. (2009). Increasing incidence of differentiated thyroid cancer in the United States, 1988–2005. *Cancer, 115*(16), 3801-3807. doi:10.1002/cncr.24416

Chen, T. T., Chute, J. P., Feigal, E., Johnson, B. E., & Simon, R. (2000). A model to select chemotherapy regimens for phase III trials for extensive-stage small-cell lung cancer. *Journal of the National Cancer Institute, 92*(19), 1601-1607. doi:10.1093/jnci/92.19.1601

Chen, Y., DeMets, D. L., & Gordon Lan, K. (2004). Increasing the sample size when the unblinded interim result is promising. *Statistics in medicine, 23*(7), 1023-1038. doi:10.1002/sim.1688

Chikere, C. C., & Nwoka, J. (2015). The systems theory of management in modern day organizations-A study of aldgate congress resort limited port harcourt. *International Journal of Scientific and Research Publications, 5*(9), 1-7.

Chow, S.-C., & Chang, M. (2008). Adaptive design methods in clinical trials–a review. *Orphanet journal of rare diseases, 3*(1), 1. doi:10.1186/1750-1172-3-11

Chow, S. C., Corey, R., & Lin, M. (2012). On the independence of data monitoring committee in adaptive design clinical trials. *Journal Biopharmaceutical Statistics, 22*(4), 853-867. doi:10.1080/10543406.2012.676536

Chuang-Stein, C. anderson, K., Gallo, P., & Collins, S. (2006). Sample size reestimation: a review and recommendations. *Drug Information Journal, 40*(4), 475-484. doi:10.1177/216847900604000413

Clinical Trials Transformation Initiative. (n.d.-a). AACT (Aggregate Analysis of ClinicalTrials.gov) database. Retrieved from https://www.ctti-clinicaltrials.org/aact-database

Clinical Trials Transformation Initiative. (n.d.-b). Download AACT (Aggregate Analysis of ClinicalTrials.gov). Retrieved from https://aact.ctti-clinicaltrials.org/download

Coffey, C. S. (Producer). (2017, March 2). Increasing the practicality of innovative adaptive trial designs [Webinar]. *In Society for Clinical Trials Webinar Series*. Retrieved from http://www.sctweb.org/members/webinar.cfm?vid=244646401

Coffey, C. S., & Kairalla, J. A. (2008). Adaptive clinical trials. *Drugs in Research & Development, 9*(4), 229-242.

Coffey, C. S., Levin, B., Clark, C., Timmerman, C., Wittes, J., Gilbert, P., & Harris, S. (2012). Overview, hurdles and future work in adaptive designs: perspectives from a National Institutes of Health-funded workshop. *Clinical Trials, 9*(6), 671-680. doi:10.1177/1740774512461859

Collinson, F. J., Gregory, W. M., McCabe, C., Howard, H., Lowe, C., Potrata, D., . . . Wah, T. (2012). The STAR trial protocol: a randomised multi-stage phase II/III study of sunitinib comparing temporary cessation with allowing continuation, at the time of maximal radiological response, in the first-line treatment of locally advanced/metastatic renal cancer. *BioMed Central Cancer, 12*(1), 598. doi:10.1186/1471-2407-12-598

Committee for Medicinal Products for Human Use [CHMP]. (2006). Reflection paper on methodological issues in confirmatory clinical trials with flexible design and

analysis plan. London: European Medicines Agency (EMEA). Retrieved from http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003616.pdf

Creswell, J. W., & Creswell, J. D. (2014). *Research design: Qualitative, quantitative and mixed methods approaches* (4th ed.). Thousand Oaks, CA: Sage Publications.

Dahlberg, S. E., Shapiro, G. I., Clark, J. W., & Johnson, B. E. (2014). Evaluation of statistical designs in phase I expansion cohorts: the Dana-Farber/Harvard Cancer Center experience. *Journal of the National Cancer Institute, 106*(7), dju163. doi:10.1093/jnci/dju163

De Bot, K., Lowie, W., & Verspoor, M. (2007). A dynamic systems theory approach to second language acquisition. *Bilingualism: Language and cognition, 10*(1), 7-21. doi:https://10.1017/S1366728906002732

DeAngelo, D., Stevenson, K., Dahlberg, S., Silverman, L., Couban, S., Supko, J., . . . Turner, A. (2015). Long-term outcome of a pediatric-inspired regimen used for adults aged 18–50 years with newly diagnosed acute lymphoblastic leukemia. *Leukemia, 29*(3), 526-534. doi:10.1038/leu.2014.229

Dimairo, M., Boote, J., Julious, S. A., Nicholl, J. P., & Todd, S. (2015). Missing steps in a staircase: a qualitative study of the perspectives of key stakeholders on the use of adaptive designs in confirmatory trials. *Trials, 16*(1), 430. doi:10.1186/s13063-015-0958-9

Doroshow, J. H., & Parchment, R. E. (2008). Oncologic phase 0 trials incorporating clinical pharmacodynamics: From concept to patient. *Clinical Cancer Research, 14*(12), 3658-3663. doi:10.1158/1078-0432.CCR-07-4562

Driscoll, J. J., & Rixe, O. (2009). Overall survival: still the gold standard: why overall survival remains the definitive end point in cancer clinical trials. *The Cancer Journal, 15*(5), 401-405. doi:10.1097/PPO.0b013e3181bdc2e0

Drug Information Association. (n.d.). Adaptive design scientific working group. Retrieved from https://communities.diaglobal.org/communities/community-home?CommunityKey=202e7955-be99-409b-8e0f-a8790f63c81b

Dwan, K., Gamble, C., Williamson, P. R., & Kirkham, J. J. (2013). Systematic review of the empirical evidence of study publication bias and outcome reporting bias—an updated review. *PLoS One, 8*(7), e66844. doi:10.1371/journal.pone.0066844

Fleming, T. R., & DeMets, D. L. (1996). Surrogate end points in clinical trials: Are we being misled? *Annals of Internal Medicine, 125*(7), 605-614. doi:10.7326/0003-4819-125-7-199610010-00011

Food and Drug Administration. (1997). The FDA Modernization Act of 1997. Retrieved from https://www.fda.gov/RegulatoryInformation/LawsEnforcedbyFDA/SignificantAmendmentstotheFDCAct/FDAMA/ucm089179.htm

Food and Drug Administration, U. S. (2004). Challenge and opportunity on the critical path to new medical products. Retrieved from

http://www.fda.gov/downloads/ScienceResearch/SpecialTopics/CriticalPathInitiative/CriticalPathOpportunitiesReports/ucm113411.pdf

Food and Drug Administration, U. S. (2006). Critical path opportunities list. *US Department of Health and Human Services Food and Drug Administration, Rockville, MD, USA*. Retrieved from http://wayback.archive-it.org/7993/20180125075636/https://www.fda.gov/downloads/ScienceResearch/SpecialTopics/CriticalPathInitiative/CriticalPathOpportunitiesReports/UCM077258.pdf

Food and Drug Administration, U. S. (2010). Guidance for industry: Adaptive design clinical trials for drugs and biologics. Retrieved from http://www.fda.gov/downloads/Drugs/.../Guidances/ucm201790.pdf

Food and Drug Administration, U. S. (2011). Guidance for industry: Clinical considerations for therapeutic cancer vaccines. Retrieved from https://www.fda.gov/downloads/biologicsbloodvaccines/guidancecomplianceregulatoryinformation/guidances/vaccines/ucm278673.pdf

Food and Drug Administration, U. S. (2017). 22 case studies where phase 2 and phase 3 trials had divergent results. Retrieved from https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=0ahUKEwjf1MvFp4TXAhUI7yYKHRHaCHEQFggoMAA&url=https%3A%2F%2Fwww.fda.gov%2Fdownloads%2FAboutFDA%2FReportsManualsForms%2FReports%2FUCM535780.pdf&usg=AOvVaw3jiZGRxcNvxvXs6ofN7KfH

Food and Drug Administration, U. S. (2018a). Adaptive Designs for Clinical Trials of Drugs and Biologics Guidance for Industry [Draft]. Retrieved from https://www.fda.gov/downloads/drugs/guidances/ucm201790.pdf

Food and Drug Administration, U. S. (2018b). Master Protocols: Efficient Clinical Trial Design Strategies to Expedite Development of Oncology Drugs and Biologics [Draft]. Retrieved from https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM621817.pdf

Food and Drug Administration, U. S. (n.d.-a). Critical path initiative. Retrieved from http://www.fda.gov/ScienceResearch/SpecialTopics/CriticalPathInitiative/ucm076689.htm

Food and Drug Administration, U. S. (n.d.-b). Understanding unapproved use of approved drugs "off label". Retrieved from https://www.fda.gov/forpatients/other/offlabel/default.htm

Gallo, P., Chuang-Stein, C., Dragalin, V., Gaydos, B., Krams, M., & Pinheiro, J. (2006). Adaptive designs in clinical drug development—an executive summary of the PhRMA working group. *J Biopharm Stat, 16*(3), 275-283. doi:10.1080/10543400600614742

Geifman, N., & Butte, A. (2016). A patient-level data meta-analysis of standard-of-care treatments from eight prostate cancer clinical trials. *Scientific data, 3*, 160027. doi:10.1038/sdata.2016.27

Gillison, M. L., Zhang, Q., Jordan, R., Xiao, W., Westra, W. H., Trotti, A., . . . Ang, K. K. (2012). Tobacco smoking and increased risk of death and progression for patients with p16-positive and p16-negative oropharyngeal cancer. *Journal of Clinical Oncology, 30*(17), 2102-2111. doi:10.1200/JCO.2011.38.4099

Gluud, L. L. (2006). Bias in clinical intervention research. *American journal of epidemiology, 163*(6), 493-501. doi:10.1093/aje/kwj069

Gravdal, K., Halvorsen, O. J., Haukaas, S. A., & Akslen, L. A. (2007). A switch from E-cadherin to N-cadherin expression indicates epithelial to mesenchymal transition and is of strong and independent importance for the progress of prostate cancer. *Clinical Cancer Research, 13*(23), 7003-7011. doi:10.1158/1078-0432

Griffiths, F., Lindenmeyer, A., Powell, J., Lowe, P., & Thorogood, M. (2006). Why are health care interventions delivered over the internet? A systematic review of the published literature. *Journal of medical Internet research, 8*(2). doi:10.2196/jmir.8.2.e10

Grol, R. (2001). Improving the quality of medical care: building bridges among professional pride, payer profit and patient satisfaction. *Journal of American Medical Association, 286*(20), 2578-2585. doi:10.1001/jama.286.20.2578

Guardia-Olmos, J., De la Fuente-Solana, E. & Lozano-Fernandez, L.M. (2008). Bayesian inference for binomial populations. Bayesian estimation for child depression prevalence. *Advances and Applications in Statistics, 9*(1), 13-35.

Hansen, A. R., Graham, D. M., Pond, G. R., & Siu, L. L. (2014). Phase 1 trial design: Is 3 + 3 the best? *Cancer Control, 21*(3), 200-208. doi:10.1177/107327481402100304

Hartford, A., Thomann, M., Chen, X., Miller, E., Bedding, A., Jorgens, S., . . . Morgan, C. (2018). Adaptive Designs: Results of 2016 Survey on Perception and Use. *Therapeutic Innovation & Regulatory Science*, 2168479018807715.

Hartung, D. M., Zarin, D. A., Guise, J.-M., McDonagh, M., Paynter, R., & Helfand, M. (2014). Reporting discrepancies between the clinicaltrials. gov results database and peer-reviewed publicationsdiscrepancies between clinicaltrials. gov and peer-reviewed publications. *Annals of Internal Medicine, 160*(7), 477-483.

Hatfield, I., Allison, A., Flight, L., Julious, S. A., & Dimairo, M. (2016). Adaptive designs undertaken in clinical research: A review of registered clinical trials. *Trials, 17*(150), 1273-1279. doi:10.1186/s13063-016-1273-9

He, W., Gallo, P., Miller, E., Jemiai, Y., Maca, J., Koury, K., . . . Lin, M. (2016). Addressing challenges and opportunities of "less well-understood" adaptive designs. *Therapeutic Innovation & Regulatory Science, 51*(1), 60-68. doi:10.1177/2168479016663265

He, W., Gallo, P., Miller, E., Jemiai, Y., Maca, J., Koury, K., . . . Lin, M. (2017). Addressing challenges and opportunities of "less well-understood" adaptive designs. *Therapeutic Innovation & Regulatory Science, 51*(1), 60-68. doi:10.1177/2168479016663265

Hopewell, S., Loudon, K., Clarke, M. J., Oxman, A. D., & Dickersin, K. (2009). Publication bias in clinical trials due to statistical significance or direction of trial results. *The Cochrane Library.* Retrieved from

https://dspace.stir.ac.uk/bitstream/1893/22314/1/Hopewell_et_al-2009-The_Cochrane_Library.pdf

Huang, J.-H., Su, Q.-M., Yang, J., Lv, Y.-H., He, Y.-C., Chen, J.-C., . . . Zheng, Q.-S. (2015). Sample sizes in dosage investigational clinical trials: a systematic evaluation. *Drug design, development and therapy, 9*, 305-312.

Huang, X., Acero, A., Hon, H.-W., & Reddy, R. (2001). *Spoken language processing: A guide to theory, algorithm and system development*. New Jersey: Prentice-Hall.

Informa. (n.d.). TrialTrove: Clinical Trial Intelligence.

Institute of Medicine (US) Forum on Drug Discovery, D. and Translation, . (2010). *Transforming Clinical Research in the United States: Challenges and Opportunities: Workshop Summary*. . Paper presented at the 3, Challenges in Clinical Research, Washington, DC.

International Committee of Medical Journal Editors. (2017). Recommendations for the conduct, reporting, editing and publication of scholarly work in medical journals. Retrieved from http://www.icmje.org/

Ioannidis, J. P., Hozo, I., & Djulbegovic, B. (2013). Optimal type I and type II error pairs when the available sample size is fixed. *Journal of clinical epidemiology, 66*(8), 903-910.e902. doi:10.1016/j.jclinepi.2013.03.002

Jardim, D. L., Schwaederle, M., Hong, D. S., & Kurzrock, R. (2016). An appraisal of drug development timelines in the era of precision oncology. *Oncotarget, 7*(33), 53037-53046.

Jemal, A., Siegel, R., Xu, J., & Ward, E. (2010). Cancer statistics, 2010. *A Cancer Journal for Clinicians, 60*(5), 277-300. doi:10.3322/caac.20073

Kairalla, J. A., Coffey, C. S., Thomann, M. A., & Muller, K. E. (2012). Adaptive trial designs: a review of barriers and opportunities. *Trials, 13*, 145. doi:10.1186/1745-6215-13-145

Kampman, E., Vrieling, A., van Duijnhoven, F. J., & Winkels, R. M. (2012). Impact of diet, body mass index and physical activity on cancer survival. *Current nutrition reports, 1*(1), 30-36. doi:10.1007/s13668-011-0004-9

Kim, A., Ueda, Y., Naka, T., & Enomoto, T. (2012). Therapeutic strategies in epithelial ovarian cancer. *Journal of experimental & clinical cancer research, 31*(1), 14. doi:10.1186/1756-9966-31-14

Knudson-Martin, C. (1994). The female voice: Applications to Bowen's family systems theory. *Journal of Marital and Family Therapy, 20*(1), 35-46. doi:10.1111/j.1752-0606.1994.tb01009.x

Korn, E. L., & Freidlin, B. (2017). Adaptive clinical trials: Advantages and disadvantages of various adaptive design elements. *Journal National Cancer Institute, 109*(6), djx013. doi:10.1093/jnci/djx013

Kroeze, W., Werkman, A., & Brug, J. (2006). A systematic review of randomized trials on the effectiveness of computer-tailored education on physical activity and dietary behaviors. *Annals of behavioral medicine, 31*(3), 205-223. doi:10.1207/s15324796abm3103_2

Le Tourneau, C., Lee, J.J., & Siu, L.L. (2009). Dose escalation methods in phase I cancer clinical trials. *Journal National Cancer Institute, 101*, 708-720. doi:10.1093/jnci/djp079

Lerner, R. M., Almerigi, J. B., Theokas, C., & Lerner, J. V. (2005). Positive youth development a view of the issues. *The journal of early adolescence, 25*(1), 10-16. doi:10.1177/0272431604272461

Lim, S., Janzer, A., Becker, A., Zimmer, A., Schüle, R., Buettner, R., & Kirfel, J. (2009). Lysine-specific demethylase 1 (LSD1) is highly expressed in ER-negative breast cancers and a biomarker predicting aggressive biology. *Carcinogenesis, 31*(3), 512-520. doi:10.1093/carcin/bgp324

Lin, M., Lee, S., Zhen, B., Scott, J., Horne, A., Solomon, G., & Russek-Cohen, E. (2016). CBER's experience with adaptive design clinical trials. *Therapeutic Innovation & Regulatory Science, 50*(2), 195-203.

Lonial, S., & Nooka, A. K. (2016). Myeloma is not a single disease. *Journal of oncology practice, 12*(4), 287-292. doi:10.1200/JOP.2016.010926

Maxwell, J. H., Kumar, B., Feng, F. Y., Worden, F. P., Lee, J. S., Eisbruch, A., . . . Teknos, T. N. (2010). Tobacco use in human papillomavirus–positive advanced oropharynx cancer patients related to increased risk of distant metastases and tumor recurrence. *Clinical Cancer Research, 16*(4), 1226-1235. doi:10.1158/1078-0432.CCR-09-2350

Mele, C., Pels, J., & Polese, F. (2010). A brief review of systems theories and their managerial applications. *Service Science, 2*(1-2), 126-135. doi:10.1287/serv.2.1_2.126

Miller, E., Gallo, P., He, W., Kammerman, L. A., Koury, K., Maca, J., . . . Woo, K. (2017). DIA's Adaptive Design Scientific Working Group (ADSWG) best practices case studies for "less well-understood" adaptive designs. *Therapeutic Innovation & Regulatory Science, 51*(1), 77-88. doi:10.1177/2168479016665434

Modlin, D. (2018). *Getting started with Bayesian analysis*. Paper presented at the SAS Global Forum 2018, Denver, Colorado. https://www.sas.com/content/dam/SAS/support/en/sas-global-forum-proceedings/2018/2909-2018.pdf

Morgan, C. C., Huyck, S., Jenkins, M., Chen, L., Bedding, A., Coffey, C. S., . . . Wathen, J. K. (2014). Adaptive design: results of 2012 survey on perception and use. *Therapeutic Innovation & Regulatory Science, 48*(4), 473-481. doi:10.1177/2168479014522468

National Cancer Institute, U. S. (2015). What is cancer? Retrieved from https://www.cancer.gov/about-cancer/understanding/what-is-cancer

National Cancer Institute, U. S. (2017). Cancer statistics. Retrieved from http://www.cancer.gov/about-cancer/what-is-cancer/statistics

National Cancer Institute, U. S. (n.d.). NCI dictionary cancer terms. Retrieved from https://www.cancer.gov/publications/dictionaries/cancer-terms?cdrid=45333

National Institute of Health, U. S. (2009). ClinicalTrials.gov review of results submission. Retrieved from https://prsinfo.clinicaltrials.gov/ResultsDetailedReviewItems.pdf

National Institute of Health, U. S. (2010). ClinicalTrials.gov review of protocol submissions. Retrieved from https://prsinfo.clinicaltrials.gov/ProtocolDetailedReviewItems.pdf

National Institute of Health, U. S. (2015a). The National Cancer Institute policy ensuring public availability of results from NCI-supported clinical trials. Retrieved from https://grants.nih.gov/grants/guide/notice-files/NOT-CA-15-011.html

National Institute of Health, U. S. (2015b). Overview of FDAAA and other trial registration policies. Retrieved from https://prsinfo.clinicaltrials.gov/trainTrainer/Overview-FDAAA-Other-Regist-Policies.pdf

National Institute of Health, U. S. (2016, 2016, June 22). Phases of clinical trials. Retrieved from https://www.cancer.gov/about-cancer/treatment/clinical-trials/what-are-trials/phases

National Institute of Health, U. S. (2017a). About the results database. Retrieved from https://clinicaltrials.gov/ct2/about-site/results

National Institute of Health, U. S. (2017b). FDAAA 801 requirements. Retrieved from https://clinicaltrials.gov/ct2/manage-recs/fdaaa

National Institute of Health, U. S. (2018). ClinicalTrials.gov background. Retrieved from https://clinicaltrials.gov/ct2/about-site/background

National Institute of Health, U. S. (n.d.). ClinicalTrials.gov - Frequently asked questions. Retrieved from https://clinicaltrials.gov/ct2/manage-recs/faq

O'Reilly, K., Mclaughlin, A. M., Beckett, W. S., & Sime, P. J. (2007). Asbestos-related lung disease. *American family physician, 75*(5), 683-688.

Ogunsakin, R. E., & Siaka, L. (2017). Bayesian inference on malignant breast cancer in Nigeria: A diagnosis of MCMC convergence. *Asian Pacific Journal of Cancer Prevention, 18*(10), 2709-2716. doi:10.22034/APJCP.2017.18.10.2709

Organisation for Economic Co-operation and Development [OECD]. (2013). Focus on health - Directorate for employment, labour and social affairs. Retrieved from https://www.oecd.org/els/health-systems/Focus-on-Health_Cancer-Care-2013.pdf

Panza, F., Solfrizzi, V., Seripa, D., Imbimbo, B. P., Lozupone, M., Santamato, A., . . . Daniele, A. (2016). Tau-based therapeutics for Alzheimer's disease: active and passive immunotherapy. *Immunotherapy, 8*(9), 1119-1134. doi:10.2217/imt-2016-0019

Parekh, A., Buckman-Garner, S., McCune, S., ONeill, R., Geanacopoulos, M., Amur, S., . . . Hills, I. (2015). Catalyzing the critical path initiative: FDA's progress in drug development activities. *Clinical Pharmacology & Therapeutics, 97*(3), 221-233. doi:10.1002/cpt.42

Patton, W., & McMahon, M. (1999). *Career development and systems theory: A new relationship*. Belmont, CA: Thomson Brooks/Cole Publishing Co.

Patton, W., & McMahon, M. (2006). The systems theory framework of career development and counseling: Connecting theory and practice. *International*

Journal for the Advancement of Counselling, 28(2), 153-166. doi:10.1007/s10447-005-9010-1

Prasad, V., & Mailankody, S. (2017). Research and development spending to bring a single cancer drug to market and revenues after approval. *Journal of American Medical Association Internal Medicine, 177*(11), 1569-1575. doi:10.1001/jamainternmed.2017.3601

Pregelj, L., Verreynne, M.-L., & Hine, D. (2015). Changes in clinical trial length. *Nature Reviews, Drug Discovery, 14*(5), 307-308. doi:10.1038/nrd4611

Purdue, M. P., Hoppin, J. A., Blair, A., Dosemeci, M., & Alavanja, M. C. (2007). Occupational exposure to organochlorine insecticides and cancer incidence in the Agricultural Health Study. *International journal of cancer, 120*(3), 642-649. doi:10.1002/ijc.22258

Quinlan, J., Gaydos, B., Maca, J., & Krams, M. (2010). Barriers and opportunities for implementation of adaptive designs in pharmaceutical product development. *Clinical Trials, 7*(2), 167-173. doi:10.1177/1740774510361542

Quinlan, J. A., & Krams, M. (2006). Implementing adaptive designs: logistical and operational considerations. *Therapeutic Innovation & Regulatory Science, 40*(4), 437-444. doi:10.1177/216847900604000409

Raftery, A. E., & Lewis, S. M. (1996). *Markov chain Monte Carlo in Practice*. New York: Chapman & Hall/CRC.

Raftery, A. E. L., S. M. (1992). One long run with diagnostics: Implementation strategies for Markov Chain Monte Carlo. *Statistical science, 7*, 493-497.

Raghav, K. P., Mahajan, S., Yao, J. C., Hobbs, B. P., Berry, D. A., Pentz, R. D., . . . Overman, M. J. (2015). From protocols to publications: A study in selective reporting of outcomes in randomized trials in oncology. *Journal Clinical Oncology, 33*(31), 3583-3590. doi:10.1200/JCO.2015.62.4148

Rasmussen, N., Lee, K., & Bero, L. (2009). Association of trial registration with the results and conclusions of published trials of new oncology drugs. *Trials, 10*, 116. doi:10.1186/1745-6215-10-116

Reitsma, D., Combest, A., Hummel, J., & Simmons, A. . (2015). We can improve oncology trials using adaptive designs. Retrieved from http://www.appliedclinicaltrialsonline.com/we-can-improve-oncology-trials-using-adaptive-designs

Rogatko, A., Schoeneck, D., Jonas, W., Tighiouart, M., Khuri, F. R., & Porter, A. (2007). Translation of innovative designs into phase I trials. *Journal of Clinical Oncology, 25*(31), 4982-4986. doi:10.1200/JCO.2007.12.1012

Ross, S. (2010). Evidence for the relationship between diet and cancer. *Experimental Oncology, 32*(3), 137-142.

Salzberg, M. (2012). First-in-human phase 1 studies in oncology: the new challenge for investigative sites. *Rambam Maimonides medical journal, 3*(2). doi:10.5041/RMMJ.1074

SAS. (n.d.-a). Assessing Markov Chain convergence. Retrieved from http://support.sas.com/documentation/cdl/en/statug/63962/HTML/default/viewer.htm - statug_introbayes_sect008.htm

SAS. (n.d.-b). SAS Version 9.4. Cary, NC: SAS Institute, Inc.

SAS. (n.d.-c). SAS/STAT 9.2 User's Guide - Bayesian analysis: Advantages and disadvantages. 2nd. Retrieved from https://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm - statug_introbayes_sect006.htm

Scher, H. I., Nasso, S. F., Rubin, E. H., & Simon, R. (2011). Adaptive clinical trial designs for simultaneous testing of matched diagnostics and therapeutics. *Clinical Cancer Research, 17*(21), 6634-6640. doi:10.1158/1078-0432.CCR-11-1105

Schmidli, H., Bretz, F., Racine, A., & Maurer, W. (2006). Confirmatory seamless phase II/III clinical trials with hypotheses selection at interim: Applications and practical considerations. *Biometrical journal, 48*(4), 635-643. doi:10.1002/bimj.200510231

Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics, 6*(2), 461-464.

Siddiqui, M., & Rajkumar, S. V. (2012). The high cost of cancer drugs and what we can do about it. *Mayo Clinic Proceedings, 87*(10), 935-943. doi:10.1016/j.mayocp.2012.07.007

Simon, R., Rubinstein, L., Arbuck, S. G., Christian, M. C., Freidlin, B., & Collins, J. (1997). Accelerated titration designs for phase I clinical trials in oncology. *Journal of the National Cancer Institute, 89*(15), 1138-1147. doi:10.1093/jnci/89.15.1138

Spiegelhalter, D., Best, N. G., Carlin, B. P., & van der Linde, A. (2003). Bayesian measures of model complexity and fit. *Quality control and applied statistics, 48*(4), 431-432. doi:10.1111/1467-9868.00353

Statistics Solutions. (n.d.). nQuery - Sample size calculator (Version 8) [Sample size calculator]. Boston, MA: Statistical Solutions Ltd. Retrieved from https://www.statsols.com/

Steinert, Y., Mann, K., Centeno, A., Dolmans, D., Spencer, J., Gelula, M., & Prideaux, D. (2006). A systematic review of faculty development initiatives designed to improve teaching effectiveness in medical education: BEME Guide No. 8. *Medical teacher, 28*(6), 497-526. doi:10.1080/01421590600902976

Stokes, M., Chen, F., & Gunes F. (2014). *An introduction to Bayesian analysis with SAS/STAT® software.* Paper presented at the SAS Global Forum, Washington, DC.

Tse, T., Williams, R. J., & Zarin, D. A. (2009). Reporting "basic results" in ClinicalTrials. gov. *Chest, 136*(1), 295-303. doi:10.1378/chest.08-3022

Vera-Ramirez, L., Ramirez-Tortosa, M. C., Sanchez-Rovira, P., Ramirez-Tortosa, C. L., Granados-Principal, S., Lorente, J. A., & Quiles, J. L. (2013). Impact of diet on breast cancer risk: a review of experimental and observational studies. *Critical reviews in food science and nutrition, 53*(1), 49-75. doi:10.1080/10408398.2010.521600

Viele, K., Berry, S., Neuenschwander, B., Amzal, B., Chen, F., Enas, N., . . . Lindborg, S. (2014). Use of historical control data for assessing treatment effects in clinical trials. *Pharmaceutical statistics, 13*(1), 41-54. doi:10.1002/pst.1589

Viele, K., & McGlothlin, A. (2017). *Adaptive methods training – Session 1 [Webinar presentation]*. Paper presented at the Biostatistical Consulting Group Training, INC/Inventiv Health, Boston, MA.

Villaruz, L. C., & Socinski, M. A. (2013). The clinical viewpoint: definitions, limitations of RECIST, practical considerations of measurement. *Clinical Cancer Research, 19*(10), 2629-2636. doi:10.1158/1078-0432.CCR-12-2935

Warren, G. W., & Cummings, K. M. (2013). *Tobacco and lung cancer: risks, trends and outcomes in patients with cancer.* Paper presented at the American Society of Clinical Oncology educational book. American Society of Clinical Oncology. Meeting.

Wassmer, G., & Dragalin, V. (2015). Designing issues in confirmatory adaptive population enrichment trials. *Journal Biopharmaceutical Statistics, 25*(4), 651-669. doi:10.1080/10543406.2014.920869

Weber, W. A. (2009). Assessing tumor response to therapy. *Journal of nuclear medicine, 50*(Suppl 1), 1S-10S. doi:10.2967/jnumed.108.057174

World Health Organization Collaboration Centre. (n.d.). ATC (Anatomical Therapeutic Code) / DDD (Defined Daily Dose) Index 2017. Retrieved from https://www.whocc.no/atc_ddd_index/

Yi, N., Kaklamani, V. G., & Pasche, B. (2011). Bayesian analysis of genetic interactions in case–control studies, with application to adiponectin genes and colorectal cancer risk. *Annals of human genetics, 75*(1), 90-104. doi:10.1111/j.1469-1809.2010.00605.x

Zarin, D. A., Tse, T., Williams, R. J., Califf, R. M., & Ide, N. C. (2011). The ClinicalTrials. gov results database—update and key issues. *New England Journal of Medicine, 364*(9), 852-860. doi:10.1056/NEJMsa1012065

Zhuang, S. H., Xiu, L., & Elsayed, Y. A. (2009). Overall survival: a gold standard in search of a surrogate: the value of progression-free survival and time to progression as end points of drug efficacy. *The Cancer Journal, 15*(5), 395-400. doi:10.1097/PPO.0b013e3181be231d

Appendix A: Definition of Acronyms

Acronym	Definition
AACR	American Association for Cancer Research
AACT	Aggregate Analysis of ClinicalTrials.gov
ADSWG	Adaptive Design Scientific Working Group
AIC	Akaike information criteria
AICC	Akaike information criteria corrected for bias
ATC	Anatomical-Therapeutic-Chemical
BIO	Biotechnology Innovation Organization
CBER	Center for Biologics Evaluation and Research
CDER	Center for Drug Evaluation and Research
CHMP	Committee for Medicinal Products for Human Use
CL	Confidence limits
CRB	Central Review Boards
CRM	Continual Reassessment Method
CT	Clinical Trials
CTTI	Clinical Trials Transformation Initiative
DECI	Design and Evaluation of Complex Interventions
Df	Degrees of freedom
DIA	Drug Information Association
DIC	Deviance Information Criterion
DLT	Dose limiting toxicities
DMC	Data monitoring committee
ESS	Effective Sample Size
FDAAA	Food and Drug Administration Amendment Act
FDA	Food and Drug Administration
FDAMA	Food and Drug Modernization Act
HER2	Human epidermal growth factor receptor 2
HHS	United States Department of Health and Human Services
HPD	Highest posterior density
IRB	Internal Review Boards
I-SPY	Investigation of Serial Studies to Predict Your Therapeutic Response with Imaging and moLecular Analysis
MCMC	Monte Carlo Markov Chain
MeSh	Medical Subject Heading
MTA	Molecularly targeted agents
MTD	Maximum tolerated dose
NCI	United States National Cancer Institute
NE	Not estimable

NIH	United States National Institute of Health
OECD	Organisation for Economic Co-operation and Development
OR	Odds ratio
OS	Overall survival
PFS	Progression free survival
PhRMA	Pharmaceutical Research and Manufacturers of America
PLM	Postfitting analysis for very general linear models
Q1	First quartile
Q3	Third quartile
RCT	Randomized clinical trials
Ref	Reference group
RP2D	Recommended phase two dose
RR	Relative risk
SOC	Standard of care
SPP	Survival post progression
SSR	Sample Size Reassessment
US	United States
WHO	World Health Organization
WHOCC	World Health Organization Collaboration Center

Appendix B: Disease Conditions – Additional Details

Disease condition	Frequency	Percent
Anal Neoplasms	11	1.35
Appendiceal Neoplasms	2	0.25
Biliary Tract Neoplasms	4	0.49
Brain Neoplasms	3	0.37
Breast Neoplasms	123	15.09
Cervical Neoplasms	5	0.61
Cholangiocarcinoma	1	0.12
Colorectal Neoplasms	98	12.02
Digestive System Neoplasms	1	0.12
Endometrial Neoplasms	3	0.37
Esophageal Neoplasms	3	0.37
Fallopian Tube Neoplasms	6	0.74
Gastrointestinal Neoplasms	10	1.23
Head and Neck Neoplasms	25	3.07
Liver Neoplasms	4	0.49
Lung Neoplasms	117	14.36
Lymphoma	1	0.12
Lymphoma, NonHodgkin	4	0.49
Melanoma	6	0.74
Mouth Neoplasms	1	0.12
Multiple Myeloma	102	12.52
Neoplasm Metastasis	2	0.25
Neoplasms	3	0.37
Neoplasms, Glandular and Epithelial	3	0.37
Neoplasms, Second Primary	3	0.37
Neoplasms, Solid	14	1.72
Nervous System Neoplasms	1	0.12
Neuroendocrine Tumors	2	0.25
Neuroendocrine Tumour	2	0.25

Disease condition	Frequency	Percent
Oropharyngeal Neoplasms	9	1.10
Ovarian Neoplasms	38	4.66
Pancreatic Neoplasms	48	5.89
Penile Neoplasms	8	0.98
Peritoneal Neoplasms	4	0.49
Prostatic Neoplasms	77	9.45
Salivary Gland Neoplasms	1	0.12
Skin Neoplasms	5	0.61
Stomach Neoplasms	18	2.21
Testicular Germ Cell Tumor	5	0.61
Thyroid Neoplasms	4	0.49
Urethral Neoplasms	2	0.25
Urinary Bladder Neoplasms	9	1.10
Urogenital Neoplasms	2	0.25
Urologic Neoplasms	2	0.25
Uterine Cervical Neoplasms	8	0.98
Uterine Neoplasms	5	0.61
Uveal Neoplasms	2	0.25
Vaginal Neoplasms	8	0.98

Appendix C: Treatment Classification – Additional Details

Treatment Classification	Frequency	Percent
Anti-Hpv-16	4	0.49
852a	4	0.49
Abemaciclib	4	0.49
Abiraterone	10	1.23
Adavosertib	1	0.12
Afatinib	2	0.25
Aflibercept	9	1.10
Aldesleukin	13	1.60
Alectinib	4	0.49
Alisertib	3	0.37
Alpharadin	3	0.37
Anastrozole	3	0.37
Androgen	1	0.12
Anti-Hpv-16	1	0.12
Apatinib	1	0.12
Apatorsen	1	0.12
Atezolizumab	7	0.86
Axitinib	1	0.12
Azacitidine	2	0.25
Azd4547	1	0.12
Azd4877	1	0.12
Bendamustine	2	0.25
Bevacizumab	13	1.60
Bicalutamide	1	0.12
Binimetinib	1	0.12
Birinapant	3	0.37
Bortezomib	14	1.72
Buparlisib	3	0.37
Burixafor	2	0.25

Treatment Classification	Frequency	Percent
Cabazitaxel	7	0.86
Cabozantinib	5	0.61
Camptothecin	6	0.74
Capecitabine	14	1.72
Capivasertib	3	0.37
Carboplatin	21	2.58
Carfilzomib	8	0.98
Carmustine	1	0.12
Carotuximab	1	0.12
Cediranib	1	0.12
Celecoxib	1	0.12
Ceritinib	2	0.25
Cetuximab	7	0.86
Cisplatin	27	3.31
Clarithromycin	1	0.12
Crizotinib	3	0.37
Custirsen	1	0.12
Cyclophosphamide	25	3.07
Cytarabine	1	0.12
Dabrafenib	1	0.12
Dacarbazine	1	0.12
Dacomitinib	2	0.25
Dactolisib	2	0.25
Daratumumab	4	0.49
Denosumab	2	0.25
Dexamethasone	19	2.33
Dkn-01	3	0.37
Docetaxel	22	2.70
Dovitinib	4	0.49

Treatment Classification	Frequency	Percent
Doxorubicin	8	0.98
Durvalumab	1	0.12
Dutasteride	1	0.12
Elotuzumab	2	0.25
Enzalutamide	9	1.10
Epacadostat	1	0.12
Eribulin	4	0.49
Erlotinib	12	1.47
Etoposide	5	0.61
Everolimus	3	0.37
Evofosfamide	2	0.25
Exemestane	2	0.25
Figitumumab	2	0.25
Filgrastim	8	0.98
Fludarabine	13	1.60
Fluorothymidine	5	0.61
Fluorouracil	20	2.45
Flutamide	1	0.12
Fosbretabulin Tromethamine	3	0.37
Fulvestrant	4	0.49
Ganetespib	3	0.37
Gedatolisib	1	0.12
Gefitinib	3	0.37
Gemcitabine	22	2.70
Glidescope	1	0.12
Glufosfamide	1	0.12
Goserelin	2	0.25
Hydroxychloroquine	1	0.12
Ibandronate	1	0.12

Treatment Classification	Frequency	Percent
Ifosfamide	2	0.25
Imalumab	1	0.12
Imatinib	1	0.12
Indusatumab Vedotin	1	0.12
Inebilizumab	1	0.12
Ipilimumab	1	0.12
Irinotecan	13	1.60
Isis 183750	1	0.12
Ixazomib	2	0.25
Krn330	2	0.25
Kw-2478	1	0.12
Lapatinib	5	0.61
Lcl161	1	0.12
Lenalidomide	13	1.60
Lenvatinib	1	0.12
Letrozole	5	0.61
Leucovorin	11	1.35
Leuprolide	6	0.74
Levocetirizine	1	0.12
Levofolinate	2	0.25
Linsitinib	1	0.12
Loratadine	1	0.12
Luminespib	1	0.12
Lutetium Lu 177 Dotatate	2	0.25
Lymphocyte Depleting Prep Regimen	1	0.12
Melatonin	1	0.12
Melitac 12.1 Peptide Vaccine	1	0.12
Melphalan	3	0.37
Mitomycin	3	0.37

Treatment Classification	Frequency	Percent
Mk2206	7	0.86
Mln0264	1	0.12
Mm-111	1	0.12
Nab-Paclitaxel	8	0.98
Nadofaragene firadenovec	1	0.12
Naproxen	1	0.12
Necitumumab	3	0.37
Neratinib	1	0.12
Nilotinib	2	0.25
Nintedanib	4	0.49
Nivolumab	5	0.61
Nsaid	1	0.12
Octreotide	2	0.25
Olaparib	4	0.49
Orteronel	3	0.37
Osimertinib	4	0.49
Other	11	1.35
Oxaliplatin	14	1.72
Paclitaxel	27	3.31
Pacritinib	1	0.12
Palbociclib	5	0.61
Panobinostat	1	0.12
Patritumab	1	0.12
Pazopanib	1	0.12
Pegfilgrastim	1	0.12
Pembrolizumab	4	0.49
Pemetrexed	5	0.61
Pertuzumab	10	1.23
Pictilisib	1	0.12

Treatment Classification	Frequency	Percent
Plerixafor	2	0.25
Pomalidomide	3	0.37
Ponatinib	1	0.12
Prednisolone	1	0.12
Prednisone	16	1.96
Psma Adc	2	0.25
Ramucirumab	7	0.86
Regorafenib	3	0.37
Relugolix	1	0.12
Ribociclib	1	0.12
Rocuronium	1	0.12
Roflumilast	1	0.12
Rosuvastatin	1	0.12
Ruxolitinib	7	0.86
Surgery	1	0.12
Satraplatin	1	0.12
Saw Palmetto	1	0.12
Selinexor	1	0.12
Selumetinib	6	0.74
Serabelisib	1	0.12
Siltuximab	1	0.12
Sodium Deoxyribonucleate	1	0.12
Sonidegib	1	0.12
Sorafenib	2	0.25
Sotrastaurin	1	0.12
Sulfasalazine	1	0.12
Sunitinib	6	0.74
Surgery	1	0.12
Talazoparib	2	0.25

Treatment Classification	Frequency	Percent
Tamoxifen	1	0.12
Tasquinimod	1	0.12
Tavokinogene Telseplasmid	1	0.12
Tecemotide	1	0.12
Tegafur/Gimeracil/oteracil	1	0.12
Temozolomide	3	0.37
Temsirolimus	2	0.25
Testosterone Cypionate	1	0.12
Thalidomide	2	0.25
Tigatuzumab	1	0.12
Tivantinib	1	0.12
Tivozanib	2	0.25
Topotecan	1	0.12
Trabectedin	1	0.12
Trametinib	4	0.49
Trastuzumab	24	2.94
Triptorelin	2	0.25
Vandetanib	2	0.25
Veliparib	5	0.61
Vemurafenib	2	0.25
Vinorelbine	3	0.37
Vismodegib	1	0.12
Visualase Thermal Therapy	1	0.12
Vorinostat	3	0.37
Young Tumor-Infiltrating Lymphocytes	5	0.61
Young Tumor-infiltrating lymphocytes	1	0.12
Zilver Stent	1	0.12
Zoledronic Acid	1	0.12
[6r] 5,10-Methylenetetrahydrofolate	1	0.12

Appendix D: Hypothesis 2 Additional Analyses

Influence on Adaptive Design Estimates in a Bivariate Analysis of Specified Covariate with Late Stage Results as Outcome

Favorable/Unfavorable			120/76		
Univariate Covariate	DF	Estimate	Standard Error (SE)	Odds Ratio (OR) and 95% Confidence Interval (CI)	Overall *P*-value/ *P*-value
Experimental treatment classification	1	NE	NE	NE	Most studies were drug trials, so limited variability for this analysis.
Cancer type					0.21
Breast	1	-0.34	0.36	0.44 (0.32,1.54)	0.34
Colorectal	1	0.34	0.49	0.87 (0.08, 2.43)	0.49
Head and neck	1	0.08	0.67	0.67 (0.13, 5.71)	0.91
Lung	1	-0.48	0.36	0.38 (0.07, 2.17)	0.19
Multiple myeloma	1	1.22	0.62	2.08 (0.27, 16.26)	0.05
Neoplasms, Solid	1	0.48	0.64	1.00 (0.13, 7.89)	0.45
Ovarian	1	0.04	0.40	0.64 (0.11, 3.76)	0.92
Pancreatic	1	-0.62	1.29	0.33 (0.01, 8.18)	0.63
Prostatic	1	-1.20	0.60	0.19 (0.03, 1.40)	0.04
Other	---	---	---	---	---
Experimental drug classification					0.32
Anthracyclines and related substances	---	NE	NE	NE	NE
Anti-androgens	---	NE	NE	NE	NE
Corticosteroids	---	NE	NE	NE	NE
Folic acid metabolite	---	NE	NE	NE	NE
Glucocorticoids	---	NE	NE	NE	NE
Monoclonal antibodies	---	NE	NE	NE	NE
Nitrogen mustard analogues	---	NE	NE	NE	NE
Other alkylating agents	---	NE	NE	NE	NE
Other antineoplastic agents	---	NE	NE	NE	NE
Other hormone antagonists and related agents	---	NE	NE	NE	NE
Other immunosuppressants	---	NE	NE	NE	NE
Platinum compounds	---	NE	NE	NE	NE
Protein kinase inhibitors	---	NE	NE	NE	NE

Influence on Adaptive Design Estimates in a Bivariate Analysis of Specified Covariate with Late Stage Results as Outcome

Favorable/Unfavorable			120/76		
Pyrimidine analogues	---	NE	NE	NE	NE
Taxanes		NE	NE	NE	NE
Other	---	---	---	---	---
Duration between studies (months)					0.03
Quartile 4	1	-0.62	0.25	0.33 (0.14, 0.78)	0.01
Quartile 3	1	0.37	0.27	0.90 (0.37, 2.2	0.17
Quartile 2	1	-0.23	0.25	0.49 (0.21, 1.15)	0.36
Quartile 1 (reference)	---	---	---	---	---
Study funding					0.25
Industry	1	-0.29	0.44	0.59 (0.32, 1.10)	0.51
NIH	1	0.05	0.83	0.83 (0.07, 9.71)	0.95
Other	---	---	---	---	---
Sensitivity Variables					
Type of endpoint					

Surrogate	---	NE	NE	NE	NE
Clinical	---	NE	NE	NE	NE
Sample size					0.48
	1	-0.003	0.004	1.00 (0.99, 1.01)	0.46
Biomarker					0.64
Yes	1	0.17	0.3556	1.40 (0.35, 5.64)	0.64
No	---	---	---	---	---

Table A4

Full Analysis of Hypothesis Two

Appendix E: Hypothesis 3 – Additional Analyses

Table A5

Full Analysis of Hypothesis Three

Appendix F: Framework Figure Permission for Reprint from Author

2/17/2019

RE: Framework for design and evaluation of complex interventions to improve health

Raymond Fitzpatrick

Fri 1/26/2018 1:41 PM

Yes, do go ahead and use the figure, Donna.
Kind regards
Ray Fitzpatrick

From: Donna Levy
Sent: 26 January 2018 14:00
To: Raymond Fitzpatrick
Cc: Donna Levy
Subject: Framework for design and evaluation of complex interventions to improve health

Hi Dr. Fitzpatrick,

I am working on my dissertation pertaining to the complexities of clinical trials and adaptive designs in oncology research, and your research framework appears to fit appropriately for a conceptual framework. I was wondering if I could get approval to use Figure 2 (Iterative view of development of randomised controlled trials of complex interventions) from your paper within my dissertation?

If you could please let me know, I would appreciate it. If you have questions or concerns, also please let me know.

Thanks.

Donna

Campbell, M., Fitzpatrick, R., Haines, A., Kinmonth, A. L., Sandercock, P., Spiegelhalter, D., & Tyrer, P. (2000). Framework for design and evaluation of complex interventions to improve health. BMJ: British Medical Journal, 321(7262), 694.