American College of Radiology

Annual Progress Report: 2012 Formula Grant

Reporting Period

July 1, 2013 – June 30, 2014

Formula Grant Overview

The American College of Radiology received $1,851,408 in formula funds for the grant award period January 1, 2013 through December 31, 2016. Accomplishments for the reporting period are described below.

Research Project 1: Project Title and Purpose

Exploration of the RTOG Clinical Trial Database – Beyond Protocol-Specified Endpoints – For over 40 years, the Radiation Therapy Oncology Group (RTOG) has been funded by the National Cancer Institute (NCI) to conduct clinical trials seeking to improve the survival and quality of life of cancer patients. Drawing upon this vast resource of demographic, treatment, and outcome data, the researchers will test new hypotheses and explore associations that were not defined in the treatment protocols for patients with gynecologic, head and neck, lung, and prostate cancers. These analyses may inform and/or lead to future protocols.

Anticipated Duration of Project

7/1/2013 – 12/31/2016

Project Overview

This project aims to analyze data that have been collected in previous RTOG studies. The specific research objectives of this project relate to six data analysis efforts.

Aim 1: Correlation of Radiation Therapy Dose Volume Histogram (DVH) Data with GI Toxicity in Post-Operative Cervical and Endometrial Cancer Patients Treated with IMRT: RTOG 0418 is a Phase II trial that evaluated the use of IMRT in post-operative cervical and endometrial cancer patients. Using data collected from this trial, we will correlate the radiation therapy DVH data, relative to the amount of bowel receiving radiation, with reported GI adverse events.

Aim 2: Evaluation of the Impact of Treatment Time for Head and Neck Cancer Patients Treated with Radiation Therapy: Using data collected from 3 RTOG Phase III Head and Neck Cancer Trials (RTOG 9003, 9111, and 9501), we will evaluate whether or not a longer radiotherapy treatment time is associated with a significantly worse clinical outcome.

Aim 3: Evaluation of Outcome in Squamous Cell Carcinoma of the Head and Neck (SCCHN) Based on

Age: Using data collected from several RTOG combined modality Head and Neck cancer trials (RTOG 8527, 9003, 9111, 9703, 9903, 9914, 0129, and 0522), we will evaluate efficacy outcomes & Adverse Events (AE) by age categorizations: ≥ 70 vs. < 70 and ≤ 60 vs. 61-69 vs. ≥ 70.

Aim 4: Evaluation of Incidental Cardiac Irradiation on Toxicity and Survival in Stage IIIA/IIIB Non-Small Cell Lung Cancer (NSCLC) Patients Treated with Chemoradiotherapy: Using data from RTOG 0617, we will correlate the radiation therapy dose volume histogram data, relative to the amount of heart receiving radiation, with cardiac and pulmonary AEs and overall survival.

Aim 5: Evaluation of Hormone Therapy Length on Outcome for Intermediate Risk Prostate Cancer Patients: We will evaluate whether or not radiotherapy with long-term hormones (28 months) is associated with better outcome than radiotherapy with short-term hormones (4 months) for the RTOG 9202 subset of intermediate-risk prostate cancer patients.

Aim 6: Evaluation of Changes in Serum Testosterone Levels in Prostate Cancer Patients Treated with Radiotherapy Alone: We will evaluate associations between radiated area (prostate vs. whole pelvis) and changes in serum testosterone for the patients treated on the radiotherapy alone arm of RTOG 9408.

Principal Investigator

Kathryn A. Winter, MS
RTOG Director, Statistics
American College of Radiology
1818 Market Street Suite 1600
Philadelphia, PA 19103

Other Participating Researchers

Wendy Seiferheld; Asha George; Chen Hu; Daniel Hunt, PhD; Jonathan Harris, MS; Jennifer Moughan, MS; Rebecca Paulus, MS; Ed Zhang, PhD; Jennifer Presley, RT – employed by American College of Radiology

Expected Research Outcomes and Benefits

Aim 1: Correlation of Radiation Therapy Dose Volume Histogram Data with GI Toxicity in Post-Operative Cervical and Endometrial Cancer Patients Treated with IMRT: Results from this aim will inform dose constraints for future IMRT GYN trials to help minimize GI adverse events.

Aim 2: Evaluation of the Impact of Treatment Time for Head and Neck Cancer Patients Treated with Radiation Therapy: Results from this aim may impact the approach to treatment breaks and will help to inform treatment time components of future trials.

Aim 3: Evaluation of Outcome in Squamous Cell Carcinoma of the Head and Neck (SCCHN) Based on Age: Results from this aim may identify subsets of patients by age that are associated with a

benefit from certain treatment regimens and/or associated with significantly better/worse treatment adverse events. This will help to form the basis for future SCCHN clinical trials.
Aim 4: Evaluation of Incidental Cardiac Irradiation on Toxicity and Survival in Stage IIIA/IIIB Non-Small Cell Lung Cancer (NSCLC) Patients Treated with Chemoradiotherapy: Results from this aim will help to define critical anatomic cardiac structures and inform the dose constraints to be used on future lung and other trials where the heart is in the area of the radiation treatment field.

Aim 5: Evaluation of Hormone Therapy Length on Outcome for Intermediate Risk Prostate Cancer Patients: Results from this aim may lead to a trial to definitely evaluate radiotherapy with long-term hormones in the intermediate-risk prostate cancer patient population.

Aim 6: Evaluation of Changes in Serum Testosterone Levels in Prostate Cancer Patients Treated with Radiotherapy Alone: Results from this aim may lead to improvements in the amount of scatter radiation to the testes. This data may also serve as a control group for a future project to evaluate associations between radiotherapy modality and serum testosterone changes.

Summary of Research Completed

Aim 1: No progress to report for this period.

Aim 2: No progress to report for this period.

Aim 3: No progress to report for this period.

Aim 4: No progress to report for this period.

Aim 5: Evaluation of Hormone Therapy Length on Outcome for Intermediate Risk Prostate Cancer Patients:
Statistical analyses were done, the results were presented at the 2013 ASTRO Annual Meeting and are summarized below.

Radiation Therapy Oncology Group (RTOG) Protocol 9202 was a randomized trial testing long-term adjuvant androgen deprivation (LTAD) after initial androgen deprivation vs initial androgen deprivation only (STAD) with external-beam radiotherapy (RT) in patients with mostly high-risk prostate cancer. Of interest, some intermediate risk patients were eligible**.** RTOG 9202 demonstrated a benefit in all study endpoints except overall survival, with the exception of the subset of patients with a Gleason score 8-10. More recently, RTOG 9408 found an overall survival advantage in patients with T1b- T2b prostate adenocarcinoma with PSA less than 20, with the bulk of the benefit observed among intermediate risk patients. Thus, while STAD was validated in 9408, it is not known whether patients in the intermediate risk subset would experience an additional survival benefit with longer duration androgen deprivation. The inclusion of some intermediate risk patients in 9202 allows the exploration of whether LTAD had any incremental benefit above STAD. For this analysis, the endpoints of overall survival (OS), disease-specific survival (DSS), and prostate specific antigen failure (PSAF) were evaluated.

An analysis was done of all patients enrolled in RTOG 9202 who were defined as intermediate-risk disease based on having T2 disease and PSA<10 and Gleason=7 OR T2 disease and PSA 10-20 and Gleason <7. This review yielded a total of 133 patients who fit this definition: 74 in the STAD arm and 59 in the LTAD arm. The Kaplan-Meier method was used to estimate the survival rate for OS, with the log-rank test used to test the significance between the two treatment arms for the endpoint of OS. The cumulative incidence approach was used to estimate the DSS and PSAF rates, with Gray's test used to test the significance between the two treatment arms for these two endpoints. Two-sided test was used at a significance level 0.05.

With over 11 years of median follow up, 39 patients were alive in the STAD group and 33 were alive in the LTAD group. There was no difference in overall survival (10-year estimates 61% STAD vs. 65% LTAD, p=0.53) between the two groups. With regards to disease-specific survival, there were a total of only 4 failures in the STAD group and 3 in the LTAD group. 10-year DSS was 96% vs. 96%, respectively, p=0.72. PSAF occurred in 38 patients in the STAD group and 33 in the LTAD group. 10-year PSAF rates were 53% and 55%, respectively, p=0.99.

LTAD was not associated with a benefit in terms of overall survival, disease-specific survival and PSA failure rates in the subset of patients analyzed in this study with intermediate risk prostate cancer. While the subset was relatively small, the treatment assignment was randomly applied and a trend in favor of longer hormonal therapy would have been of interest. Given the small number of disease-specific deaths observed and the lack of a benefit with respect to any of our endpoints, this analysis does not suggest that exploration of longer duration hormonal therapy is worth testing in the intermediate-risk subset of prostate cancer patients.

Aim 6: Evaluation of Changes in Serum Testosterone Levels in Prostate Cancer Patients Treated with Radiotherapy Alone:
Statistical analyses were done, the results were presented at the 2013 ASTRO Annual Meeting and are summarized below.

In light of studies suggesting that radiotherapy (RT) may influence serum testosterone (ST) levels for patients treated for localized prostate cancer, reviewed data on testosterone changes for patients treated with RT alone on the Phase III trial, RTOG 9408, was reviewed**.**

Patients enrolled on RTOG 9408 (clinical tumor stage: T1b-T2b, prostate-specific antigen (PSA) <20ng/ml) were randomized between RT alone and RT plus 4 months of total androgen ablation. RT consisted of either whole pelvic radiotherapy to 46.8Gy plus a 19.8Gy prostate boost for a total dose of 66.6Gy (WPRT) or treatment to the prostate alone for a total dose of 68.4Gy (PORT). Most patients received WPRT. Only patients with the lowest risk features (PSA <10ng/ml and Gleason score ≤5 or a negative lymph node dissection) were assigned to receive PORT. RT was delivered at 1.8Gy per fraction. For this analysis, serum testosterone levels were investigated at the following collection periods: at study enrollment; completion of RT; and at first follow-up 3 months after completion of RT. The Wilcoxon signed rank test was used to compare change in pre and post treatment serum testosterone levels in patients who were randomized to the RT alone arm.

A total of 2028 patients were enrolled on RTOG 9408 and 992 were randomized to receive RT alone. Of those 992 patients, 904 (91%) had baseline serum testosterone values available and completed RT. Of these 904, immediate and 3 month post RT testosterone levels were available for 768 and 553 respectively. Excluding 10 patients who received hormonal therapy off protocol, 766 and 543 respectively, were analyzed. Pre-treatment median testosterone level for all patients was 370ng/dL, (5^{th} percentile value of 167 and a 95^{th} percentile value of 926) which did not differ significantly between the WPRT and PORT groups. For the entire group, at completion of RT the median, 5^{th} percentile, and 95^{th} percentile serum testosterone change values were -26ng/dL, -310, and 360 respectively ($p<0.01$). At the 3 month follow up, the median, 5^{th} percentile, and 95^{th} percentile serum testosterone changes values were -33ng/dL, -300, and 238, respectively ($p<0.01$). The median change in serum testosterone showed a statistically significant trend, indicating a decrease in serum testosterone level from baseline.

For evaluable patients treated with WPRT at end of RT (n=679), the median, 5^{th} percentile, and 95^{th} percentile serum testosterone change values were -29ng/dL, -365, and 452, respectively ($p<0.01$); and for WPRT patients at 3-month follow-up (n=476), the median, 5^{th} percentile, and 95^{th} percentile serum testosterone change values were -33ng/dL, -313, and 245, respectively ($p<0.01$). Patients treated with PORT at end of RT (n=87), the median, 5^{th} percentile, and 95^{th} percentile serum testosterone change values were -12ng/dL, -202, and 101, respectively ($p=0.01$); and for PORT patients at the 3-month follow-up (n=67), the median, 5^{th} percentile, and 95^{th} percentile serum testosterone change values -37ng/dL, -240, and 110, respectively ($p<0.01$).

Radiation therapy for prostate cancer, as delivered on RTOG 9408, was associated with statistically significant change in serum testosterone values, indicating a decline. This may be secondary to scatter radiation to the Leydig cells. There is no evidence that these changes in serum testosterone have any direct impact on PSA control rates or on post treatment quality of life.

Research Project 2: Project Title and Purpose

Community Learning of a Prediction Model for Treatment Outcome in Head and Neck Cancer Patients for Radiation Therapy Decision Support – Personalized medicine for head and neck cancer (HNC) is promising, but validated decision support systems are needed to make the promise a reality. A decision support system relies on a model to predict treatment outcome (e.g. survival, quality of life, toxicity). Such a model can be developed through a machine learning process using a well-organized database and query system that is designed for a community based rapid learning approach. This project aims to build such a model to guide head and neck radiotherapy treatment, and includes the development of an IT infrastructure for the Radiation Therapy Oncology Group (RTOG) to manage and deploy the clinical trial data needed for machine learning and building predictive models for radiotherapy treatment of HNC.

Anticipated Duration of Project

7/1/2013 – 12/31/2016

Project Overview

This project will test the hypothesis that it is feasible to build a decision support system to provide personalized radiotherapy treatment plans for head and neck cancer (HNC) patients. Three specific aims are proposed as follows:

Specific Aim 1. Build an IT infrastructure for machine learning. Clinical trial data used for machine learning requires full semantic interoperability so that the local data can be translated into a centralized database. The IT infrastructure also needs to support a community based rapid learning approach where routine patient data from many institutions in many countries is shared for learning. The design of the underlying technology will combine a local semantic interoperable environment with a distributed learning framework. When new patients (or new members) in the community become available an updated model can be learned.

Specific Aim 2. Modeling of survival in HNC based on our previous study. Utilizing an established machine learning system, a model that predicts the treatment outcome (including survival, toxicity, etc.) in HNC patients will be studied using the RTOG protocol 0522 clinical trial data. Classical approaches such as the logistic regression model as well as the so-called second-generation machine learning approaches such as Bayesian networks will be employed for modelling. The model performance is quantitatively evaluated.

Specific Aim 3. Extend the model by including more predictive parameters to improve model performance. Functional imaging procedures are employed more widely in cancer diagnosis and treatment. Large amounts of biological and molecular information become available as well with the advancement of sequencing technology. The project will explore these additional predictors in modeling to enhance the predictive performance of models.

Principal Investigator

Ying Xiao, PhD
Radiation Oncology Core Lab Physicist
Jefferson Medical College
G-321D Bodine Center
111 South 11th Street
Philadelphia, PA 19107

Other Participating Researchers

James Galvin, PhD – Consultant
Elizabeth O'Meara; Ed Zhang, PhD; Jonathan Harris, MS – employed by American College of Radiology
Yunfeng Cui, PhD; Jialu Yu, PhD; Yutao Gong, PhD – employed by Jefferson Medical College
Jiazhou Wang, MS – employed by Fudan University Shanghai Cancer Center – currently at Thomas Jefferson University

Expected Research Outcomes and Benefits

Extracting knowledge in the form of a prediction model can be used to change care delivery. Very specific questions like "what radiation dose should this head and neck cancer patient receive for an expected survival of X% at two years" can be answered. These are the type of questions that are being posed at the point of care.

The predictive models are built from a machine learning system that learns and shares knowledge while leaving the data behind the firewalls of the institutions. This system will be established as we complete Specific Aim 1 of this project. The important next step is to prove the rapid learning hypothesis that knowledge can be extracted from coordinated databases of routine care and clinical trial data sources. Using this system, learning can be done without data leaving the institute that holds the data.

The machine learning infrastructure can also be used to study other types of disease. Once deployed, the system can be leveraged in multiple research projects targeted at specific treatment modalities and specific cancers. Also, the technology is such that it can easily be applied outside of cancer. An open source solution, using semantic web technology and machine learning techniques, will boost the use of rapid learning in health care in the United States. The predictive models based on machine learning will be used to provide decision support in the personalized medicine era to give patients the best outcome: longer survival and better quality of life.

Summary of Research Completed

<u>Specific Aim 1. Build an IT infrastructure for machine learning.</u>

While building the infrastructure, the team developed a manuscript detailing the completed work. The abstract of the paper is as follows.

Abstract 1: Validation of a rectal cancer survival and local control model in routine clinical data
Purpose: The risk of local recurrence, metastases and overall survival of locally advanced rectal cancer after preoperative chemoradiation and curative surgery can be estimated by the prediction nomograms which have been validated in European clinical trial populations. This study is to validate these nomograms from Europe to classify the risk of survival for locally advanced rectal cancer in a Chinese cohort.
Methods and Materials: From 2006 to 2012, clinical data of 277 consecutive locally advanced rectal adenocacimoma treated with preoperative chemoradiation and curative surgery from Shanghai Cancer Center were retrospectively collected and used for external validation. Concordance index (C-index) and calibration plot were used to assess the performance of these nomograms in our population.
Results: The C-index for these published nomograms was 0.68, 0.83 and 0.70 in predicting local recurrence, distant metastases and overall survival in the Chinese population, respectively. Kaplan-Meier curves indicated good discriminating performance of local recurrence; however, it wasn't successful in discriminating low-risk and medium-risk groups in distant metastases and overall survival. Calibration plots showed an underestimation of local recurrence, distant metastases and overall survival between predicted and observed probabilities at 5 years for relapse and survival.

Conclusions: We externally validated these three nomograms in predicting 5-year LR, DM and OS of locally advanced rectal cancer patients after preoperative chemoradiation and curative surgery with good discrimination in a single Chinese cohort. Further validation in other routine clinical databases is necessary and a renewed prediction model with more data from different countries and more specific prognostic factors will enhance the application possibility for the personalized treatment of locally advanced rectal cancer.

Specific Aim 2. Modeling of survival in HNC based on our previous study.

A manuscript detailing our research for this aim is in preparation. The abstract is below.
Abstract 2: A feasibility study on incorporating clinical trial quality assurance parameters into outcome prediction in head and neck radiotherapy treatment
Purpose: To investigate the impact of radiation treatment clinical trial quality assurance (RTQA) on treatment outcome in a phase III trial for advanced head and neck cancer with development of a predictive model incorporating RTQA parameters.
Methods: RTQA for RTOG 0522 included initial institution credentialing of RT technologies and individual RT case reviews. The case review processes (including contour and dosimetry evaluations) were performed by radiation oncologist and radiation physicist co-chairs. RTQA grades (per-protocol, variation acceptable and deviation unacceptable) are given to contouring of target volume (TV), organ at risk (OAR) and dose-volume coverage of targets as defined in the protocol. The relationship between RTQA parameters and treatment outcome are analyzed with predictive modeling. A logistic regression model is established that includes RTQA parameters, age, T-stage, equivalent dose in fractions of 2 Gy (EQD2), tumor location, and hemoglobin levels. This model is compared to the one without incorporating RTQA parameters. The model prediction accuracy is validated by cross-validation and C-statistic methods.
Results: The contour (TV and OAR) quality grades did not correlate with two-year overall survival. The target dose-volume quality grade is slightly related to overall survival ($p=0.094$), and the correlation test shows that target dose-volume quality grade is an independent predictive factor. The model incorporating RTQA parameters shows that the two-year survival ratio is 86.2%, 83.6% and 80.6% for the three scores of target dose-volume quality of per-protocol, variation acceptable and deviation unacceptable. The area under the curve (AUC) indicator is 0.737 and 0.733 for the overall survival model with RTQA parameters and without RTQA parameters.
Conclusion: The results demonstrate that it is feasible to incorporate RTQA parameters into outcome modeling. This capability is of critical importance for evaluating RTQA methods as related to outcome in head and neck cancer.

Specific Aim 3. Extend the model by including more predictive parameters to improve model performance.

There was no progress to report.

Research Project 3: Project Title and Purpose

Discovery of Plasma Biomarkers of Doxorubicin and Trastuzumab Induced Cardiotoxicity in Breast Cancer – The overall objective of this proposal is to discover novel circulating biomarkers using powerful proteomic profiling methods to identify patients at increased risk for doxorubicin and trastuzumab-induced cardiotoxicity, before conventional decreases in ejection fraction or heart failure are evident. The key deliverables from this study are: 1) we will identify specific protein biomarkers indicative of early, subclinical cardiotoxicity; 2) we will gain insight into the mechanisms of doxorubicin and trastuzumab-induced cardiotoxicity, leading to new targeted therapies to prevent and treat this disease; and 3) we will build a multi-disciplinary collaboration for the study of cardiotoxicity biomarkers that we can expand to other cancer therapies.

Anticipated Duration of Project

1/1/2013 – 12/31/2016

Project Overview

Doxorubicin and trastuzumab (Herceptin®) are used widely in the treatment of breast cancer, are highly effective, and have led to important survival gains. However, these agents carry a substantial risk of cardiovascular morbidity and mortality. There is currently no adequate methodology to recognize patients at high risk for cardiac complications, prior to overt disease. The overall objective of this proposal is to discover novel circulating biomarkers using powerful proteomic profiling methods to identify patients at increased risk for both doxorubicin and trastuzumab-induced cardiotoxicity. Basic studies suggest potential mechanisms for cardiac dysfunction include oxidative stress, altered neuregulin/ErbB signaling, and anti-angiogenesis,[1-3] but the true relevance of these findings in humans and the precise mechanisms of cardiotoxicity remain to be elucidated. Furthermore, doxorubicin and trastuzumab cardiotoxicity are likely secondary to multiple altered and potentially differing pathways, and not one single mechanism. The broad working hypothesis of this proposal is that multiple circulating biomarkers, identified through discovery proteomics, will detect cancer therapy-induced cardiotoxicity in patients before conventional decreases in Left Ventricular Ejection Fraction (LVEF) or heart failure (HF) are evident. In breast cancer patients undergoing therapy with doxorubicin and trastuzumab, we will determine if patterns of change over time in protein markers differ between patients who experience cardiotoxicity and those who do not. In Aim 1, we will identify novel plasma biomarkers associated with cardiotoxicity in breast cancer patients treated with doxorubicin and trastuzumab. In Aim 2, we will verify the most promising biomarkers associated with doxorubicin and trastuzumab cardiotoxicity.

Principal Investigator

Bonnie Ky, MD, MSCE
Assistant Professor of Medicine and Epidemiology
University of Pennsylvania School of Medicine
Translational Research Center, 11-105

Philadelphia, PA 19104

Other Participating Researchers

David Speicher, PhD – employed by Wistar Institute

Expected Research Outcomes and Benefits

The key outcomes from this novel study that will advance the field of cardio-oncology and improve the overall cardiovascular and oncology care of a growing cancer population are as follows: we will determine the utility of discovery plasma proteomics in identifying patients at risk for doxorubicin and trastuzumab-induced cardiotoxicity; and we will identify specific protein biomarkers whose changes in abundance levels are indicative of the early-stage development of cardiotoxicity. These two results alone will substantially advance the field of cancer therapy cardiotoxicity risk prediction. We will gain specific insights into the mechanisms of cancer therapy induced cardiotoxicity which has the potential to lead to new targeted therapies to prevent and treat cardiotoxicity. This strong foundation of research has the potential to grow into multiple additional studies : 1) further verification and validation of the biomarkers identified herein; 2) pursuit of biologic mechanism leads; 3) development of new cardioprotective therapies indicated by the biologic leads; and 4) expansion of biomarker discovery and validation to additional cancers and cancer therapies. This work will serve as a critical launching pad to further build a cardio-oncology translational research program statewide and nationally, and strengthen collaborations between investigators at the University of Pennsylvania, Eastern Cooperative Oncology Group (ECOG), and American College of Radiology Imaging Network (ACRIN). It is anticipated, pending funding from other sources, a working group will be convened, comprised of cardiologists, oncologists, academic clinicians, and researchers, with the goals of developing strategies to enhance the detection of cardiotoxicity; innovative and cost-effective strategies for treatment and follow-up of cardiotoxicity; and recommendations for the management of cardiac comorbidities in cancer survivors. Health Research Funds from the Department of Health shall not be used to pay for expenses related to the work of this Committee.

Summary of Research Completed

Background

Doxorubicin and trastuzumab are two commonly used breast cancer therapies that have led to significant improvements in cancer survival, but carry a significant risk of cardiotoxicity. This risk of cardiotoxicity differs according to the use of these agents singly or in combination, as well as according to the dosing sequence. As such, there are three relevant treatment groups that carry a substantive risk of cardiotoxicity: 1) Doxorubicin, without trastuzumab (Dox only), 2) Trastuzumab followed by doxorubicin (Trastuzumab-Dox), and 3) Doxorubicin followed by trastuzumab (Dox-Trastuzumab). In this project, we will focus our proteomics discovery experiments on cases and matched controls from the first two groups, and discovery studies of the third group are being conducted by the Speicher laboratory under separate pilot funding. For the proteomics discovery efforts during the past year, we have made progress in two areas, specifically, the careful selection and intensive preparation of Dox only case and control plasma

samples, and major improvements in the depth of analysis and throughput of the proteome discovery methods.

Case/Control Selection

Three Dox only cardiotoxicity cases and matched controls were selected from an ongoing, independent longitudinal prospective cohort study defining the cardiotoxic effects of doxorubicin and/or trastuzumab. Patients were enrolled prior to doxorubicin and/or trastuzumab therapy, with standardized blood, echocardiography, and clinical data collection at prespecified intervals. Cases were defined as patients who had suffered from a decline in echocardiography core-lab quantitated left ventricular ejection fraction of ≥10% to <50%, consistent with established definitions. Furthermore, all cases were prescribed cardiac medications for treatment of their cardiomyopathy, representative of the most significant cases of cardiac dysfunction. Controls were matched based upon age, race, nodal status, and hormonal status to minimize the confounding effects of age/race as well as cancer severity. Furthermore, controls did not demonstrate any evidence of cardiac dysfunction as defined by clinical standards including the lack of changes in left ventricular ejection fraction and heart failure symptoms. Multiple plasma samples were selected for each case and control, in order to define the changes over time in the proteome between cases and controls. In total, these 30 samples consisted of the following timepoints: prior to any chemotherapy, during chemotherapy, at the time of cardiotoxicity diagnosis, and after the cardiotoxicity diagnosis. These specific plasma samples and changes in left ventricular ejection fraction indicative of the time course of cardiotoxicity that were selected for proteome analysis are illustrated in *Figure 1*.

Sample Processing

Processing of aliquots of all 30 selected plasma samples was initiated using the 3-D plasma proteome analysis strategy developed by the Speicher laboratory as outlined in *Figure 2.* Case and Control 1 samples were completely processed to the point where they are ready for LC-MS/MS analysis. Specifically, these samples were depleted of 20 abundant serum proteins using a ProteoPrep20 Immunodepletion Column (Sigma-Aldrich). 80 μL of plasma was filtered through a 0.22 μm microcentrifuge filter and injected onto the column. The flow-through fractions containing unbound proteins were collected, pooled, and precipitated with nine volumes of 200 proof ethanol, prechilled to -20^{o}C. Ethanol supernatants were carefully removed and protein pellets were frozen and stored at -20^{o}C. Fractions containing affinity-bound abundant proteins were collected and pooled, neutralized with 1 M NaOH, and frozen. Prior to 1-D SDS-PAGE, frozen protein pellets from ethanol precipitation of depleted plasma were thawed briefly, and resuspended in SDS sample buffer. For each sample, aliquots representing 10 μL of original plasma per lane were loaded onto 1-D SDS-PAGE mini-gels and separated until the tracking dye had migrated 2 cm. Gels were stained with Colloidal Coomassie Blue, and each lane was subsequently sliced into 20 uniform 1 mm slices. Corresponding slices from three lanes for each depleted plasma sample were combined in single wells of a 96-well pierced plate and gel slices were digested using trypsin. Furthermore, plasma samples from the other two cases and controls have been immunodepleted and stored as ethanol precipitates in preparation for multi-plexed analysis using isobaric tags (see below). Also, an additional aliquot of Case and Control 1 plasma samples were immunodepleted and stored as ethanol precipitates because the initial immunodepleted fractions had been dissolved in a Tris-SDS buffer that was not compatible with

the new isobaric tag approach (see *Figure 3*). These milestones were targeted in our project application.

Mass Spectrometer Analyses Strategy and Optimization

However, before committing the large amount of mass spectrometer time required to analyze these samples, we reevaluated our analysis strategy in light of the acquisition in December 2013 of a new mass spectrometer, a Thermo Electron Q Exactive Plus, by the Speicher laboratory. Initial benchmark tests with standard samples showed that the new instrument, as expected, was far more sensitive, had a faster duty cycle, and yielded higher resolution MS/MS scans that improved peptide identification. These features enabled greater depth of analysis of several standard samples. To determine whether the new instrument would substantially affect plasma proteome analyses, a pilot test of immunodepleted human plasma fractions was conducted. These data showed that the combination of the new instrument and an optimized HPLC gradient resulted in identification of approximately twice as many peptides and twice as many unique proteins from the same set of plasma proteome fractions.

This improvement was considered to be particularly critical for the current study because of the complexity of plasma, with protein concentrations ranging more than 10 orders of magnitude. Most tissue damage- specific proteins are present at very low levels in blood, and are therefore very difficult to detect. By improving the depth of analysis, we can greatly expand the detection of low abundance proteins, which should improve the likelihood that we will detect a greater number and more robust cardiac toxicity biomarkers. Specifically, it is likely that the best biomarkers will be those that are present at the lowest abundance levels, i.e., those proteins that are specifically shed by cardiac or endothelial tissue in response to early stage injury or stress.

For these reasons, it is highly desirable to analyze our samples using this new instrument. Unfortunately, due to the high performance of the instrument, it has been in heavy demand with the highest priority access given to those projects that contributed to the purchase of the instrument. Furthermore, because our current proteome analysis pipeline uses label-free quantitation based upon MS signal intensities of peptides (*Figure 2*), it is important that all samples to be compared are analyzed on the instrument contiguously. Hence, this project not only requires extensive instrument time, but this time must be in a single block. As shown in the left panel of *Figure 4*, a total of 1350 hrs (57 days) of instrument time are needed for the entire Dox only discovery experiments. Accounting for instrument calibration, instrument repair and the inevitable re-analyses of samples where instrument problems were encountered, it is expected that these analyses would require close to 3 months of continuous access to the instrument. As this is not practical, we have explored alternative strategies that could provide the depth of analysis that this new instrument promises.

Isobaric Tagging of Samples

Fortunately, an additional feature of the Q Exactive Plus compared with our older OrbiTrap XL instruments is that this instrument has an efficient high energy collision cell for high energy collisional dissociation (HCD) of peptides. This enables the use of isobaric tags such as Thermo's tandem mass tags (TMT) or AB Sciex's iTRAQ tags. These reagents covalently tag samples after the trypsin digestion. Multiple samples can then be mixed and analyzed in a single set of LC-MS/MS runs. The mixing of multiple samples does not increase the complexity of the

peptide mixtures being analyzed because all versions of the tag have the same precursor mass (MS1 scan). However, HCD partially cleaves the tag into a reporter and balance group. Because the reporter ion for each sample is different, quantitative comparisons of reporter ion intensities are feasible for detected peptides across all samples that were differentially tagged and pooled.

The most extensive multiplexing available is the TMT 10-plex kit. By using one or more reference samples, we can compare yields of all detected peptides (and corresponding proteins) across all plasma samples for a given therapeutic regimen. For example, for the Dox only discovery experiments, we will use two reference samples; a pool of all case plasma samples and a pool of all control plasma samples. These two references plus eight individual plasma samples will be included in each set of TMT samples. Hence, all 30 individual plasma samples can be analyzed in a total of four TMT experiments (Figure 4).

Use of the TMT tags requires an overall redesign of the experimental workflow as well as another fractionation method as shown in *Figure 4.* Reference and experimental samples are mixed after chemically tagging individual tryptic digests. Therefore, any variations in sample preparation prior to trypsin digestion will contribute noise to the quantitative comparisons. Hence all sample fractionation other than the initial immunodepletion should be performed at the peptide level. For this reason, the tryptic digests will be fractionated into at least 20 fractions using high pH reverse separation prior to LC-MS/MS. The Speicher lab has previously shown that high pH separation yields better depth of analysis compared with gel slices or other fractionation methods. We therefore expect that 20 high pH fractions analyzed on the new instrument will yield approximately twice as many protein identifications as use of the current 3-D method with the OrbiTrap XL mass spectrometer.

The Speicher lab is currently setting up the TMT method, and as noted above, frozen immunodepleted pellets of plasma from three Dox only and matching control samples have been prepared and are ready for the TMT labeling. All Dox only samples will be analyzed by LC-MS/MS, as outlined in *Figure 4*, in this next reporting period. This will only require about 7.5 days of instrument time, which is realistic.

Figure 1. *Plots of ejection fraction for Doxorubicin Cases and Controls*. Red arrows indicate the point of clinical diagnosis of cardiotoxicity for each Case. Plasma blood draws that are being used in the proteome-based biomarker discovery studies are indicated by squares for cases and triangles for controls.

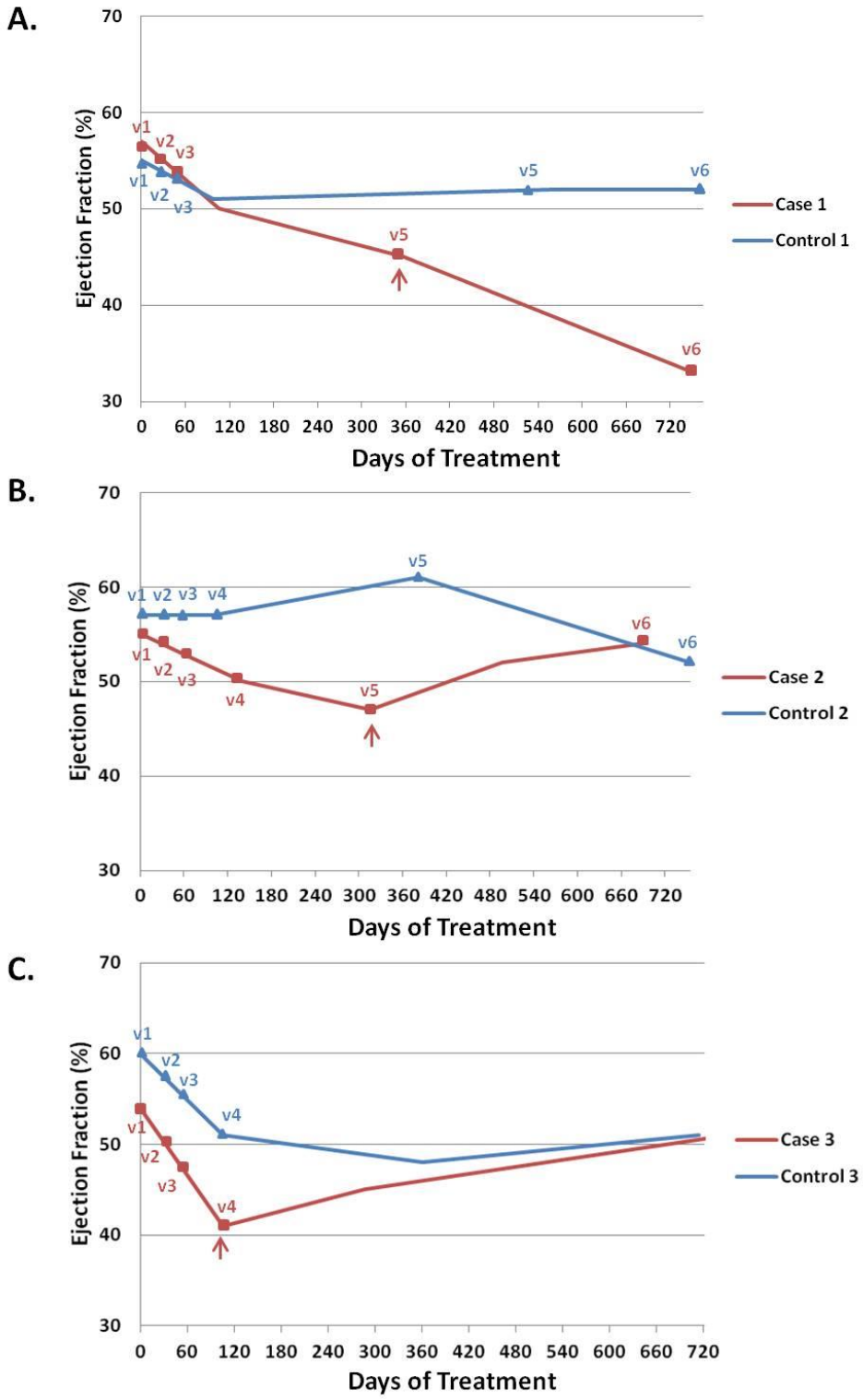


Figure 2. *Current strategy for 3-D label-free quantitative discovery of candidate biomarkers of cardiotoxicity*. Immunodepletion and 1-D SDS-PAGE followed by LC-MS/MS are used to compare patients diagnosed with cardiotoxicity and matched controls. MaxQuant label-free analysis software and other bioinformatic tools are then used to identify quantitative changes between the two groups.

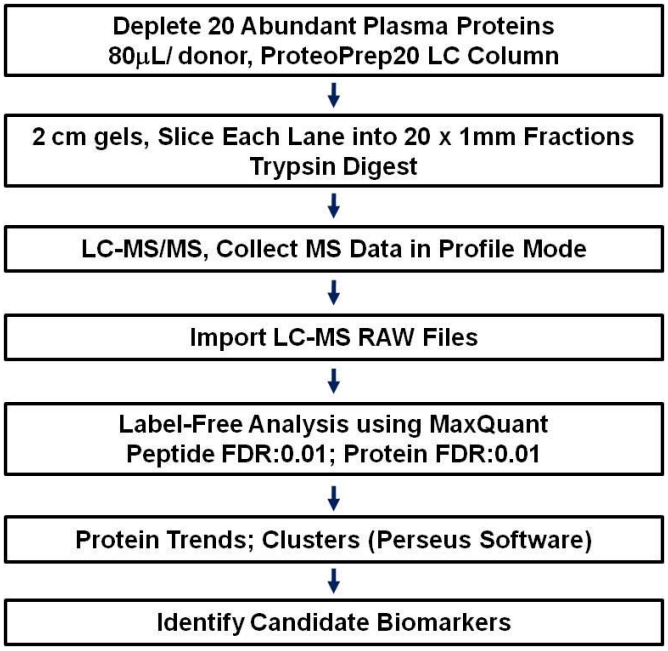


Figure 3. *Representative immunodepletion chromatogram of a patient plasma sample using the ProteoPrep 20-LC column.* Low-abundance (depleted) plasma proteins were pooled and concentrated for downstream analysis on 1D gels and LC-MS/MS. High- abundance (bound) proteins were neutralized and stored at -20°C for possible future use.

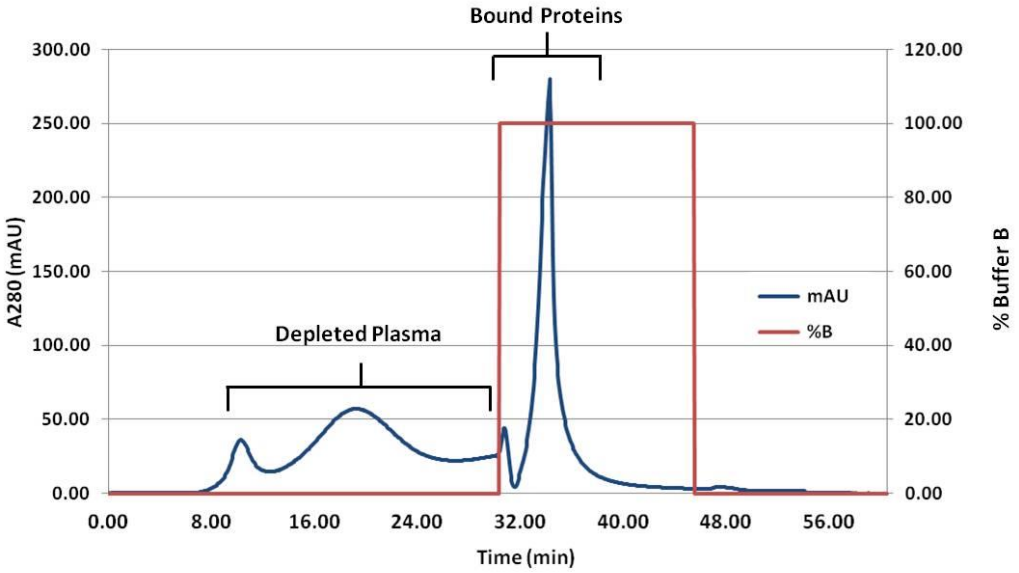
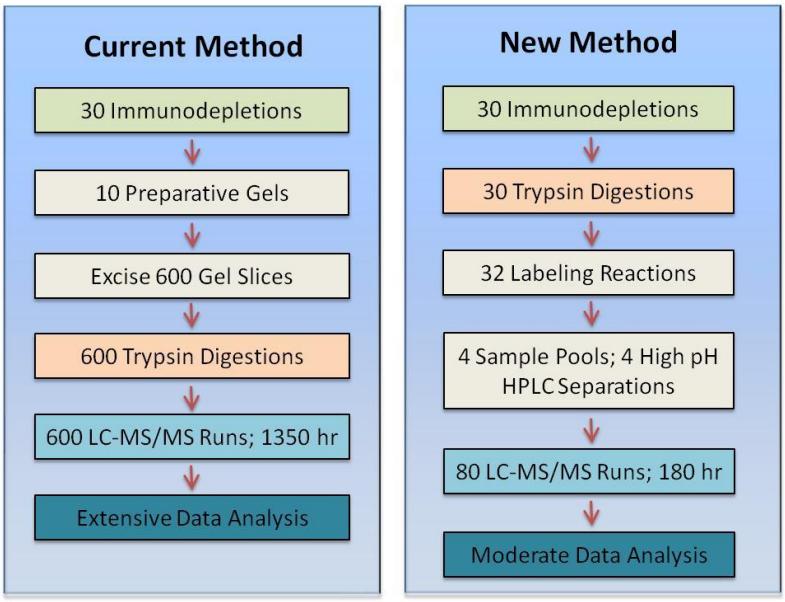


Figure 4. *Comparison of our current plasma biomarker discovery method and the new higher throughput method using isobaric tags.* The workload for analysis of three Dox only cases and matching controls is illustrated. Corresponding time intensive steps in the two methods are highlighted with the same color. Throughput will be greatly increased with the new method primarily because the number of tryptic digests is greatly reduced and total mass spectrometer time, which is the greatest bottleneck, is reduced from 1350 hours to 180 hours. Additional benefits are: 1) quantitation will be more precise and will extend across all cases and controls, and depth of analysis will be increased, thereby allowing detection of lower abundance biomarkers.



Research Project 4: Project Title and Purpose

Novel Statistical Analysis and Evaluation Methods for Multiple Endpoints in Cancer Clinical Trials – Clinical trials provide critical evidence necessary to advance clinical development in cancer research. The increasing number of promising new interventions mandates the improvement in clinical trial design and analysis, such that we can a) better understand disease progression; b) address clinical interests more quickly and efficiently; and c) conserve and optimize resources by terminating unpromising trials early. To address these needs, we propose a series of methodological projects aimed at addressing current questions in multiple endpoints in cancer clinical trials. These projects encompass a range of needs and challenges that apply broadly to cancer clinical trials and clinical research in general.

Anticipated Duration of Project

1/1/2013 – 12/31/2016

Project Overview

Aim 1: *Assessment of correlation between PFS and OS based on a Weibull model*: Progression-

free survival (PFS) has been used as a surrogate marker for overall survival (OS) in oncology clinical trials. Accurate estimation of correlations between the two endpoints is important for trial design and outcome modeling. In previous work, an exponential model was considered for this purpose. However, observed hazard rates are often non-constant across time. In this research we aim to establish a Weibull correlation model which can provide more realistic estimates for this important quantity.

Aim 2: *Estimating Hazard of Failure Over Time in Early Prostate Cancer* – Typically, time to event data is summarized using aggregate measures such as time to event functions (i.e., survival curves, cumulative incidence curves) or cumulative hazards. The hazard function, being a dynamic time-varying process, may reveal features of the failure pattern over time that may have both biologic and clinical implications. However, hazard functions present challenges in terms of estimation and interpretation. In prostate cancer specifically, there are numerous questions regarding an individual's risk of failures of different types (biochemical failure, frank clinical disease, prostate cancer death and competing cause death) that have bearing on clinical management, as well as on gaining a better understanding of the disease process. We will investigate and compare different recently developed hazard estimation methods, and apply these to practical questions in long-term follow-up after treatment for localized prostate cancer.

Aim 3: *Evaluation of PFS and OS based on a progressive multistate model*: In oncology clinical trials, PFS is often considered as a putative surrogate endpoint for OS due to its clinical relevance and correlation with OS. However, the high correlation between PFS and OS, as well as the improvement in PFS alone do not always translate into an improvement in OS directly, therefore systematic evaluation and appropriate statistical models for PFS and OS are needed to address this issue. We propose to investigate and identify factors that may influence statistical inference of OS with reference to that of PFS.

Principal Investigator

Chen Hu, PhD
Senior Statistician
American College of Radiology
1818 Market St. Suite 1600
Philadelphia, PA, 19103

Other Participating Researchers

James J. Dignam, PhD; Qiang (Ed) Zhang, PhD – employed by American College of Radiology;
Alex Tsodikov, PhD – employed by University of Michigan, Ann Arbor, MI
Vanja Dukic, PhD – employed by University of Colorado, Boulder, CO
Yimei Li, PhD – employed by Children's Hospital of Philadelphia

Expected Research Outcomes and Benefits

Clinical trials are a critical component of cancer research to advance effective interventions to prolong the survival of patients and save lives, and it is only through systematic and

comprehensive evaluation in a clinical trial setting that the risks and benefits of treatment options can be assessed. However, the process of cancer clinical research can be slower than expected and resources are limited, especially with the tremendous amount of information that needs to be collected. Meanwhile, for clarity and robustness, a single primary trial endpoint (outcome of interest) must be chosen. To improve the process of cancer clinical research, a better understanding of multiple types of failure endpoints (disease recurrence of different types, death from cancer, death from other causes, etc.) experienced after cancer treatment is needed. This may offer additional pivotal insights into treatment efficacy, as well as inform trial design and analysis. These observations also provide information on disease natural history over time. A more efficient analysis and treatment evaluation strategy making use of all this information could improve both knowledge acquisition and patient care, which may rely heavily on our knowledge of the relationship between the multiple endpoints observed sequentially in cancer clinical trials. We propose three areas of statistical methodology research that have immediate practical implications for cancer clinical trials. These novel statistical methods can more directly assess risks, benefits and effects on investigative therapeutic agents, and will increase the trial operational efficiency and produce more informative outcomes. These investigations will provide a concrete demonstration of the worth of these innovative concepts and advance knowledge in cancer research and treatment.

Summary of Research Completed

Aim 1: *Assessment of correlation between PFS and OS based on a Weibull model*

During the past year, we have worked on drafting the manuscript after including a third clinical trial example. The third study, RTOG 9111 is a randomized trial that compared three treatments for patients with locally advanced cancer of the larynx. The failure event of PFS was defined as laryngectomy or all-cause death. In all three trials the overall survival was not significantly different among treatment arms and therefore, in our following analyses, we simply included patients in all treatment arms. All distribution parameters are estimated using the tested codes and the correlations calculated using the newly derived five theorems under each model. Figure 1.1 shows the estimated Kaplan-Meier (KM) curve for OS together with the predicted OS based on Theorem 5 from both the Exponential model and Weibull model, for RTOG 9111. We can see that both predicted curves show some departures from the observed KM curve but the predicted curve from Weibull model is closer to the KM curve suggesting a better fit with Weibull model.

Table 1.1 presents the estimated parameters from Exponential model and Weibull model. In the first example, the estimated $\log(\lambda)$ in the Weibull model is very close to 0, so that the Weibull model essentially reduced to the Exponential model. Therefore the estimated $\log(\lambda)$s under the two models are almost identical. In the second example, the estimated $\log(\lambda)$ in the Weibull model is larger than 0, and therefore the estimated $\log(\lambda)$s under the two models are somewhat different. In the third example, the estimated $\log(\lambda)$ in the Weibull model is smaller than 0, and again leading to different estimates of $\log(\lambda)$s under the two models.

Table 1.2 presents the estimated correlations from the Exponential model and the Weibull model, by substituting MLEs into Theorem 2-4. As expected, in the first example the estimated

correlations under the two models are very close to each other. In the second and third example, the correlations estimated from the Weibull model are somewhat larger than those from the Exponential model.

Aim 2: *Estimating Hazard of Failure Over Time in Early Prostate Cancer*

Initial work on this project involves selection and preparation of data sets from the RTOG prostate cancer trials portfolio that are most useful in both extending hazard estimation methods and providing the basis for addressing key clinical questions. We selected for initial work the landmark RTOG 9202 trial, which compared short term vs. long term androgen deprivation therapy (ADT) in high risk localized disease. This large trial with mature follow-up over 12 years provides ideal data for evaluating patterns of biochemical recurrence, distant metastasis, and death over time in men treated by what has become a standard intervention (ADT of some duration after radiation). There is particular interest in how men who underwent different durations of ADT initially (with longer being more favorable in the trial) are experiencing recurrence events in later years of follow-up. The data was prepared, documented, and shared with collaborating colleagues in Colorado in order to begin this project. Initial analyses confirmed this dataset as useful for both extending hazard estimation methods and examining long-term patterns of recurrence and mortality.

Work on the development and testing of improved methods to estimate the hazard function was then begun, with particular focus on obtaining stable estimates at later points in follow-up time, when the number of failure events and patients at risk may be small and thus the estimate becomes very unstable. As expected, the RTOG 9202 data was found to have periods where few failures are observed. For example, only 13% of the subjects were observed to have biochemical failure after 4.9 years (the median time to biochemical failure), and only 1.5% of subjects were observed to have biochemical failure after 10 years. Even for all-cause mortality, an event that increases monotonically from diagnosis time, there are intervals of relatively sparse event counts that could lead to unstable estimates.

Briefly, the multiresolution hazard (MRH) estimator (described in detail in the application and in Bouman, Meng, Dignam, Dukic. JASA 2007) is a recursive tree-based estimator that partitions the time axis into equal size disjoint intervals or 'bins' and estimates the hazard in piecewise fashion within bins. Specifically, we approximate the hazard rate with a set of corresponding hazard increments, d_j , j = 1, . . . , J, where each d_j represents the aggregated hazard rate over the j^{th} time interval, ranging from (t_{j-1},t_j). We assume that $J = 2^M$, where $M > 0$ and can be chosen in a variety of ways; for example, using model selection criteria or clinical insights. Modeling J in this manner allows for flexible estimation of the hazard rate, as its estimate can be adjusted based on the desired value of M. The cumulative hazard H equals the sum of the 2^M hazard increments d_j , j=1, . . . 2^M . We then define "split parameters" $R_{m,p} = H_{m,2p}/H_{m-1,p}$, m = 1,2,...,M − 1, p = 1,...,2^m − 1, where $H_{m,p}$ is a specific resolution-level hazard recursively defined such that $H_{m,p} \equiv H_{m+1,2p} + H_{m+1,2p+1}$. These split parameters determine the shape of the hazard rate, are between 0 and 1, and are useful for initial parameterization required for estimation. The d_j in the final time resolution are estimated using the split parameters and the estimate of the cumulative hazard. All parameters in the model are estimated using Markov chain Monte Carlo (MCMC).

The choice of the level of the maximal resolution in the MRH prior is driven by a compromise between the desire for detail and the amount of data: as the resolution increases (and the number of time bins increases), counts within each bin will decrease. While useful for revealing detailed patterns, large number of bins (and consequently, large number of model parameters) will require longer computing times. Similarly, more bins will eventually mean lower event count per bin, and this lower information content will translate into lower efficiency. It would thus be advantageous to devise an algorithm that could adaptively choose the appropriate number of bins over different time intervals, with an increased number of bins in the regions of high event counts, and fewer bins where the counts are low. To improve the MRH estimator addressing sparseness in event counts, the idea of 'pruning' was developed.

Pruning allows for objective means of combining of adjacent time bins with fewer observed failures, which increases the efficiency of estimates and reduces unrealistic variation in the hazard estimate. In addition, the pruning method allows us to decrease the parameter dimensions *a priori* via the pruning rule, even for a specification with a large number of bins, thus decreasing the length of time required for the MCMC estimation routine to execute. The key requirement is establishment that estimation under this reduced dimensionality does not result in loss of resolution or bias with respect to important variations in the hazard function.

After the preparation and review of the data as described above, derivation, implementation, and testing of the pruned MRH (PMRH) approach was undertaken. In brief, the algorithm begins by starting with the full MRH tree, and merges adjacent bins that are constructed via the same split parameter, $R_{m,p}$, if the estimated hazard increments in these two bins ($H_{m+1,2p}$ and $H_{m+1,2p+1}$) are statistically similar. The estimate of the hazard increment in a bin is derived as the number of observed failures within the bin divided by the number of patients at risk at the initial time point of the bin. Then, for a given level m (for m = 1,...,M), the null hypothesis $H_0 : R_{m,p} = 0.5$ is tested versus the alternative H_a: $R_{m,p} = 0.5$ (with a pre-set type I error α) for each split parameter $R_{m,p}$ ($p = 0, ..., 2_{m-1} - 1$). If the null hypothesis is not rejected, that split $R_{m,p}$ is set to 0.5 and the adjacent hazard increments are considered equal and the bins declared "fused". The hypothesis testing can be applied to all M levels of the tree or just a subset of the tree. A modified Fisher's exact test is used to test the null hypothesis, based on the 2 x 2 table composed of the number of failures within the bin time interval and at-risk patients at the end of the bin time interval for each pair of adjacent bins sharing a split parameter. This test accounts for the dependence of counts across bins, and the possible small event counts in bins at the bottom levels. It is important to note that Fisher's exact test provides a simple approximate solution to a fundamentally more complex inference problem, and that other tests and modifications can be applied for different circumstances.

Mathematical properties of the PMRH model were derived for the manuscript underway. Briefly, the estimator still retains its resolution-invariance under aggregation. Although pruning is expected to reduce sensitivity of the MRH method in identifying subtle changes in the hazard rate, the pruning method carries several advantages that may be worth considering. With pruning, the posterior hazard increment estimates are expected to be less variable compared to the equivalent non-pruned model. With some split parameters preset to 0.5, fewer total number of model parameters need to be estimated, and, consequently, the estimation time should be reduced as well.

To evaluate performance of the estimator, a simulation study was undertaken. We simulated 200 datasets closely resembling a real clinical trial (Fisher et al. 1989, 1996, Dukic and Dignam 2007; Dignam et al. 2009). In this data, the estimated hazard rate exhibited a mixture of features (bumps and flat regions), which was ideal for evaluating the PMRH method's performance. Each dataset consisted of either 200 or 1000 patients (equal numbers in treatment and control arm), and for each patient the failure time was simulated depending only on the treatment indicator covariate. We estimated a PMRH model with 5 levels (M = 5), with 32 equal length bins over the total time horizon. For each set of data, we implemented 4 different PMRH strategies: NPM4: 4-level model without any pruning, NPM5: 5-level model without any pruning, PM52: 5-level model with 4th and 5th level subject to pruning, and PM55: 5-level model with all levels subject to pruning.

All the simulations were run on a supercomputer with 1368 nodes, each containing two hex-core 2.8Ghz Intel Westmere processors with 12 cores per node and 2GB of RAM per core. For a dataset of size 200, it took about 3 hours for model PM55 to complete 1 million iterations; 5 hours for model PM52; 2.5 hours for model NPM4 and 8 hours for model NPM5. The model NPM5 takes about 2.67 times longer than model PM55, as expected: PMRH method can reduce computing time substantially for a given maximum resolution of the model.

Figure 2.1 depicts the square root of MSE of each hazard increment posterior mean in the four PMRH strategies, for simulations with 1000 patients (top) and 200 patients (bottom). PM55 model seems to have the smallest root MSE, while the NPM5 model has the largest, on average over all bins. The first few PM55 hazard increment estimators have larger root MSE than the other increments, which is due to low counts in those first few bins in our simulated data, and the estimators of hazard increments in these bins are expected to be more variable than the rest. In particular, PM55 performs poorly in those first few bins as these bins are often merged into a single bin under PM55, as the null hypothesis is rarely rejected for bins with low counts. However, if we examine the square root of the integrated MSE for hazard increments (over all bins), we see that for the 200-patient simulation the PM55 model has the smallest square root of the integrated MSE, followed by PM52 model, NPM4 model and NPM5 model. In the 1000-patient simulation, PM52 has the smallest square root of integrated MSE, followed by NPM4 model, PM55 model and NPM5 model. The difference among root-integrated MSEs among models is much smaller in the 1000 patients per-set case than that of 200 patients per-set case. As the pruning only affects the prior, and the prior effect dampens as the dataset size increases, we can expect that the estimates will be very close among the different models in larger datasets.

Figure 2.2 show the 95% probability intervals of posterior means for the 32 individual hazard increments, for the four PMRH strategies, in datasets with 1000 patients (top two plots) and with 200 patients (bottom 2 plots). The left column represents the raw PMRH results while the right column shows smoothed versions using polynomials of degree 7. While more aggressive pruning will generally result in more variation in bins with fewer counts (for example, the first and last bins), it will also tend to reduce the variability over the other bins, resulting in less variable hazard rate estimator.

Two manuscripts are in progress: one that contains the more technical aspects of the pruning extension to the MRH estimator, more details on simulations, and an example focusing on

patterns in all-cause mortality in the prostate cancer data; and the other that applies the new algorithm to modeling biochemical failure hazard in the same cohort. The former manuscript has received initial peer review and is being prepared for resubmission, and the latter will be submitted imminently.

Aim 3: *Evaluation of PFS and OS based on a progressive multistate model*

Preliminary work on this project has been performed, which involves selection and evaluation of data sets from the RTOG lung cancer and malignant glioma clinical trials. Preparations on the statistical methodology have also been initiated; these include literature review and preliminary mathematical derivations on the appropriate utility metrics to evaluate the proposed method.

Table 1.1 Estimated parameters from Exponential model and Weibull model

	Exponential			Weibull		
	log(_1)	log(_2)	log(_3)	log(_)	log(_1)	log(_2)	log(_3)
RTOG 0214	-0.846	-2.418	0.037	-0.057	-0.817	-2.382	0.043
RTOG 9413	-2.038	-3.531	-2.531	0.178	-2.273	-3.766	-2.679
RTOG 9111	-1.710	-2.768	-1.086	-0.260	-1.463	-2.524	-0.907

Table 1.2: Estimated correlations using Theorems 2-4

	Exponential			Weibull		
		Conditional			Conditional	
	PFS&OS	TTP&OS	TTP&OS	PFS&OS	TTP&OS	TTP&OS
RTOG 0214	0.897	0.679		0.895	0.901	0.682	0.897
RTOG 9413	0.452	0.233		0.446	0.509	0.269	0.515
RTOG 9111	0.820	0.496		0.812	0.835	0.511	0.814

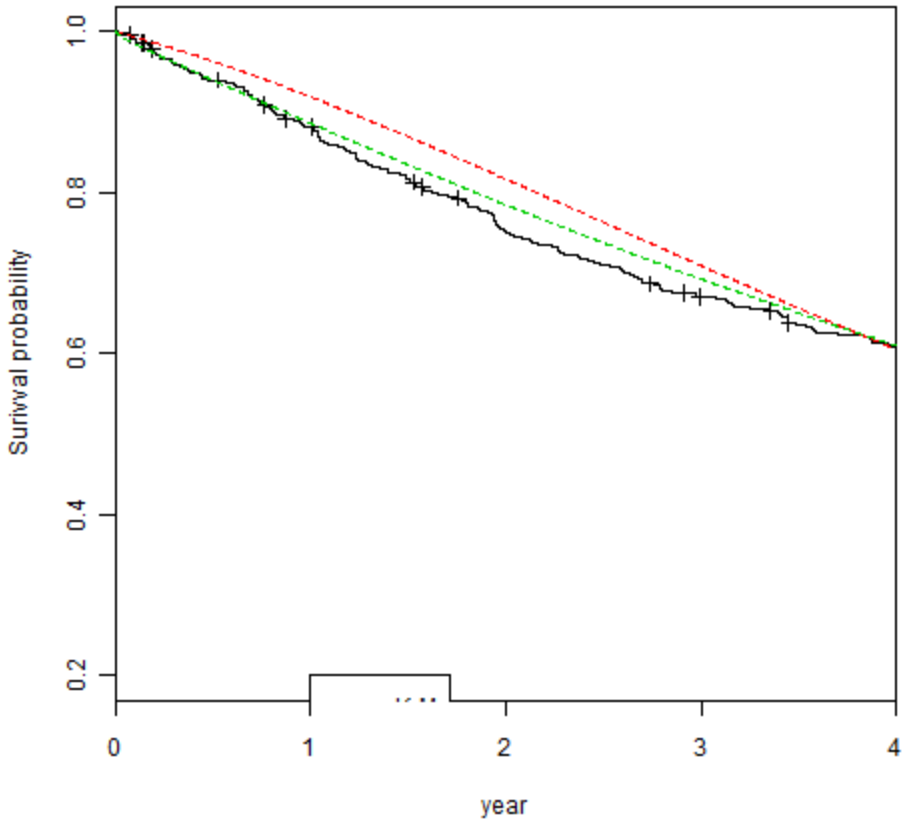


Figure 1.1: Comparison between Observed and Predicted OS Rates for RTOG 9111.

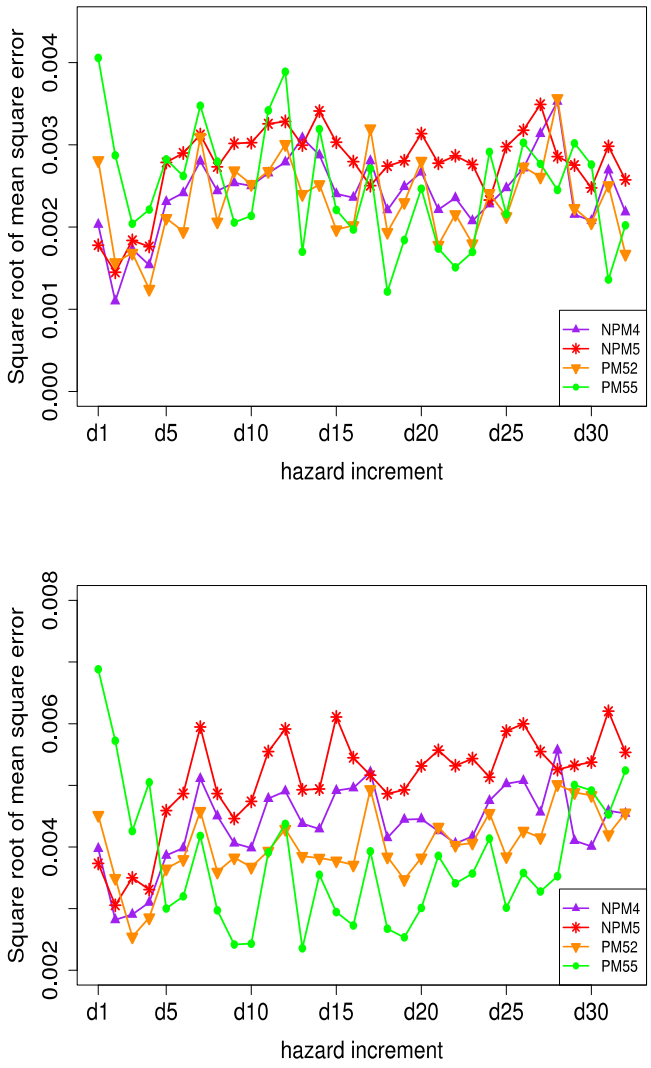


Figure 2.1: Estimated square root of the MSE for 4 PMRH hazard increment estimators (posterior means), based on 200 datasets with 1000 patients per dataset (top figure) and 200 patients per dataset (bottom figure).

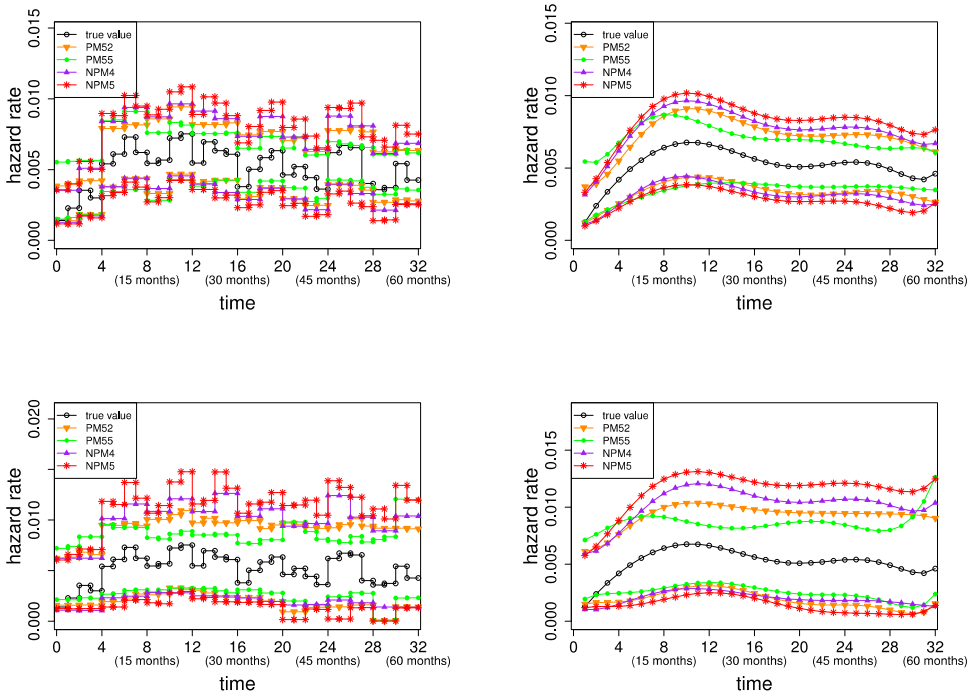


Figure 2.2: 95% probability intervals of individual hazard rate posterior means (left) and their smoothed version (right), for 4 different PMRH estimators. The top row shows the performance over 200 simulated datasets with 1000 patients per dataset, and the bottom row shows the performance over 200 simulated datasets with 200 patients each. (Model NPM5 has the widest 95% probability intervals.)