Molecular Markers of Vitamin D and Esophageal Adenocarcinoma Prognosis

Molecular Markers of Vitamin D and Esophageal Adenocarcinoma Prognosis

Elizabeth Loehrer

A Dissertation Submitted to the Faculty of

The Harvard T.H. Chan School of Public Health

in Partial Fulfillment of the Requirements

for the Degree of *Doctor of Science*

in the Department of Environmental Health

Harvard University

Boston, Massachusetts.

May 2018

Dissertation Advisor: Dr. David Christiani Elizabeth Loehrer

Molecular Markers of Vitamin D and Esophageal Adenocarcinoma Prognosis

Abstract

Esophageal cancer is the ninth leading cause of cancer and the fifth leading cause of cancer death globally, with two predominant histologies: squamous cell carcinoma and adenocarcinoma. In the last few decades, esophageal adenocarcinoma (EAC) has become the predominant histology in the U.S. Despite some improvements in treatment, EAC still has a five-year survival of less than 20%. Moreover, EAC has few prognostic markers and few if any modifiable prognostic factors. Vitamin D is one potential factor that has garnered attention for its role in tumor suppressive effects and its potential to be modified. To understand the role of vitamin D in stalling cancer progression, we also need to study underlying physiologic and metabolic factors of the vitamin D pathway.

Therefore, we utilized an ongoing survivor study of esophageal cancer patients from Massachusetts General Hospital to study the association between circulating serum levels of 25(OH)D, genetic variants in candidate genes in the vitamin D pathway, and body mass index with overall survival in EAC patients. We recruited patients around their time of diagnosis, and collected serum and whole blood from patients for biomarker and genetic analyses. Patients completed a baseline questionnaire about demographic and covariate information. Medical records were used to obtain clinical variables such as treatment regimen, pathology, and outcome information. We used extended cox models to estimate adjusted hazard ratios in all studies, adjusting for relevant confounders, modeling surgery as a time dependent covariate, and stratifying the baseline hazards by clinical stage.

We found circulating levels of 25(OH)D were not associated with overall survival in EAC, and the relationship was not modified by stage or by body mass index. Genetic variants in

the *CYP24A1* and *CYP27B1* genes were marginally associated with overall survival, but were not significant after multiple testing. Overweight and obese patients at diagnosis had reduced hazards of death, whereas patients who lost substantial weight leading up to diagnosis had significantly increased hazard of death, independent of BMI at diagnosis. Moreover, the association between change in BMI and overall survival was modified by average adult weight, with substantial weight loss being worse for patients who were in the healthy to underweight range as adults (≤27.5 kg/m2).

The findings in this dissertation help guide our understanding of the physiologic and molecular differences that might drive different clinical courses in EAC.

Table of Contents

List of Figures with Captions

List of Tables with Captions

Acknowledgements

I would like to thank my dissertation research committee, specifically my thesis advisor and committee chair, Dr. David Christiani, Dr. Rebecca Betensky, who provided guidance on the statistical analysis in my thesis, and Dr. Edward Giovannucci for his guidance on nutrition and epidemiology. All of my committee members graciously spent time in committee meetings and one on one meetings to discuss and enhance my project, and I am indebted to their time and insights.

I would like to thank the Harvard Education and Research Center (ERC) for their financial support through the ERC training grant T42 OH008416 for the duration of my doctoral studies.

I would like to thank the entire Christiani Lab for their guidance and support in helping me complete this dissertation. Li Su assisted me at every stage of the required lab work for my project to retrieve samples, plate and aliquot samples, and to ship samples to be analyzed. She also offered encouragement and support (and homemade food) as I conducted my analyses. Jongeun Rhee helped welcome me to lab with training, and then became a close friend and my work partner on projects in the lab, in our respective dissertations, and through student organization work at the school. Zhihui Wang worked side by side with me in the lab, and she would talk through analytic and coding steps for my project. Dr. Zhaozhong (ZZ) Zhu helped train me when I entered the Christiani lab and provided much guidance and advice as I developed and implemented my dissertation project. Dr. Ruyang Zhang provided excellent insights on bioinformatics and biostatistics, and was a great guide to help shape my genetic association studies in addition to being a good friend. I would like to think Lucy Ann Principe for helping orient me to the esophageal cancer study data set, and for her continued logistical and

moral support throughout my project. Andrea Shafer, Shawn Cho, and Kristof Vajda for their hard work and help recruiting patients and assisting with medical record retrieval and survival data for my project. Kevin Wong and Phil Steininger, who served as jack of all trades in the lab and helped keep everyone organized and sane. I also want to specifically thank Dr. Paula Tejera-Alvarez, Dr. Nancy Diao, Shangzhi Gao, Dr. Mohammad Rahman, Yichen Guo, Sipeng Shen, Cheng-Kuan Lin, Dr. Pi-l (Debby) Lin, Qianyu Yuan, and Xinan Wang for their help and advice with my projects and the joy and friendship that they brought to the lab.

I thank the lab of Dr. Bruce Hollis at the Medical University of South Carolina (MUSC) for their work in quantifying serum levels of 25(OH)D from our patient samples. I thank Dana-Farber/Harvard Cancer Center in Boston, MA, for the use of the Genotyping and Genetics for Population Sciences Core, which provided TaqMan OpenArray genotyping service.

I would additionally like to specially thank Dr. Marc Weisskopf and Dr. Lorelei Mucci for serving on my doctoral dissertation oral exam committee and for their early insights and advice around my dissertation project.

I would like to think Donald Halstead and Joyce LaTulippe and the terrific members of the Doctoral Dissertation Writing Seminar, who spent time editing and commenting on my manuscripts and provided invaluable advice for my dissertation writing.

My dissertation work was made possible in part by the generous funding I received from the Harvard Education and Research Center (ERC) training grant for Occupational Safety and Health (T42 OH008416). I want to thank Pat McGaffigan, Tiffany Sarkissian, and Ann Backus for their support of my project and their great work with the ERC center and programming at the Harvard Chan School.

I would like to thank the Department of Environmental Health at the Harvard T.H. Chan School of Public Health for accepting me as a doctoral student four years ago, and for the rigorous and broad training I received while in the department. I would like to give a special thanks to Dr. Barbara Zuckerman, who provided a wealth of administrative guidance and moral support throughout my time in the program.

During my doctoral degree, I was not just conducting research toward my dissertation, but I was also participating in a wide-range of external activities that helped shape my teaching and leadership skills. I would like to give special thanks for Dr. Enrique Ciffuentes, Dr. Doug Dockery, and Dr. John Evans for the opportunity to serve as a teaching assistant in EH251 Public Health in Megacities: The Environmental Dimension in Mexico City, and I would like to think Dr. Greg Wagner and Dr. David Wegman for the opportunity to serve as a teaching assistant in EH236 Epidemiology of Occupational & Environmental Health Regulations. These were my two favorite classes to take as a student at the school, and I was honored to help the teaching staff with the classes for future cohorts.

I would like to thank Jongeun Rhee and Vy Nguyen for the opportunity to serve alongside you as student officers in the Harvard Chan Environmental Justice Student Organization, and to Dr. Tamarra James-Todd, Dr. Gary Adamkiewicz, and Dr. Diana Ceballos for their support as faculty advisors to the student group. I would additionally like to thank Christine Mitchell for the opportunity to work with her to organize Weekly Wednesday Teach-Ins with local community organizations doing excellent work and advocacy about ongoing public health issues in our community. In turn, I would like to thank Leah Kane, Katherine Worthington, and Joel-Michael Martin in the Office of Student Affairs for their ongoing, fierce support of students and student organizations at the school.

I would like to thank the innumerable, incredible peers that I have had the honor to work and study with and learn from during my time at the school. I have made so many friends and learned so much from all of you. Special thanks go to Rosie Martinez, my first friend at Harvard, the person who was with me in almost every class, my go-to study partner, cheerleader, and my expert assistant on coding and technological gadgets. I also want to give special thanks to Kalé Kponee for sharing with me her wealth of knowledge, her laser perception, her unique perspectives, and her friendship.

Finally, I would like to thank my friends outside of the school, my parents: Dr. Patrick and Debi Loehrer, my late grandfather O. Les Nell, my brothers P.J. Loehrer and Dr. Andrew Loehrer, his wife and my sister-in-law Dr. Saranya Loehrer, my mother-in-law Ana Guicherit, my father-in-law Eric Guicherit and his partner Birgit Forster, my sister-in-law Jessica Guicherit and her husband and my-brother-in law Niels Eilering for their support and encouragement throughout this process. I would like to thank Gobi Loehrer Guicherit and Zuri Loehrer for the overwhelming joy they add to each day. Most importantly, I would like to thank my husband, Carlo Guicherit for his unyielding support both tangible and intangible, his devotion to making me laugh and relax, and for the adjustments he made and work he did to help me reach this goal.

Introduction

Globally, esophageal cancer is the eighth most common cancer, and the sixth leading cause of cancer death.[1] Although esophageal cancer only accounts for 1.1% of the cancer incidence in the U.S., it accounts for 2.6% of cancer death, which is even in more striking in men, accounting for 1.6% of cancer incidence but 3.9% of cancer death.[1] Esophageal cancer has two primary histologies: esophageal squamous cell carcinoma and esophageal adenocarcinoma (EAC). Most of the esophageal cancer cases that occur globally are squamous cell carcinomas.[2] However, in last two decades adenocarcinoma has become the predominant subtype in the U.S., due to a substantial decline in incidence of squamous cell carcinoma and an approximate 500% increase in adenocarcinoma incidence, particularly, a striking increase among white men.[2-8] Despite its sizeable impact on the population, esophageal cancer remains one of the least studied cancers.

Men have four times greater risk of developing EAC compared to women.[1, 2, 5, 9] In the U.S., EAC typically develops later in life (age>55years), and currently the strongest known risk factors are gastroesophageal reflux disease (GERD), Barrett's esophagus (BE), obesity, smoking, male sex, and white race, though all have low specificity for EAC.[3] Alcohol, which is a very strong risk factor for squamous cell carcinoma, may also be a risk factor for EAC.[10, 11] Diets high in fruits and vegetables and Helicobacter pylori (*H. pylori*), a gram-negative bacteria, infection seem to be protective against EAC.[3, 9]

Despite some improvements in diagnosis and treatment, esophageal cancer still has an average five-year survival around 18.4% (age, sex, and race adjusted), and the median survival is less than 1 year.[3, 9] Currently, the best prognostic markers for EAC are early stage diagnosis, tumor size, and performance status (a marker of overall health).[3] Still, even those who are

diagnosed with early stage tumors (~20%) have a five-year survival averaging 41%.[2, 9] Attempts to screen patients with GERD and BE, who have higher risk of EAC, only captures about 15% of the incident cases each year.[3] Since early detection is still a large hurdle in EAC, there is a high demand for both modifiable factors that can improve EAC patients' prognosis and biomarkers of progression that can guide targeted approaches to improving treatment.

Evidence is mounting that the vitamin D pathway, crucial for normal physiological function, plays a critical role in the pathogenesis and progression of several cancers.[12] Vitamin D is absorbed either through skin exposure to ultraviolet B radiation from sunlight (D3) or through diet and supplements (D2), which are both stored in adipocytes or converted to 25(OH)D in the liver.[13] Clinically, 25(OH)D is used to determine vitamin D levels.[13] 25(OH)D is transported throughout the body, where it is converted to $1,25(OH)_2D_3$, considered the more active metabolite. Within cells, $1,25(OH)_2D_3$ forms a complex with vitamin D receptor (*VDR*) and retinoid X receptor alpha (*RXRA*), which is referred to as the vitamin D complex.[13]

In *in vitro* study of cancer cells, $1,25(OH)_2D_3$ complex has been implicated in the role of directly and indirectly effecting expression of tumor suppressive and oncogenic proteins, inducing signaling pathways and promoting cell adhesion to limit proliferation, and inducing apoptosis.[13] In turn, *in vivo* mouse studies have shown that $1,25(OH)_2D_3$ inhibits angiogenesis and inflammatory effects in tumor cells.[13]

The vitamin D complex's role in regulating transcription of important tumor-suppressor genes is the key component to the proposed biological mechanisms by which the vitamin D pathway is involved in protective prognostics impacts. One might expect that increased vitamin D intake through diet and sunlight would always then mean increased $1,25(OH)_2D_3$ at a cellular level to ramp up the vitamin D complex's regulatory activity. Yet this is not always what we see.

Dysfunctional and normal regulation of the proteins involved in vitamin D metabolism, transport, or binding complexes may modify the associations between vitamin D intake and the association with slower progression of disease and improved overall survival. Physiologic factors like obesity can modify the metabolism of vitamin D in the body.[14] Due to adipocyte reservoirs and generally poorer metabolic function, obese subjects have less circulating 25(OH)D compared to normal weight subjects with the same sun and dietary exposure.[14] Moreover, unique tumor biology across tumor sites may mean that the vitamin D regulated genes play less of a roll in tumor suppression depending on the driving factors in tumors at different sites. Tumor stage at diagnosis, aside from unique tumor site biology, may also modify the associations between circulating vitamin D and survival. Previous work from our lab on circulating 25(OH)D serum levels in non-small cell lung cancer showed that higher levels of vitamin D improved overall survival among patients with stage Ib and IIb but not amongst patients with Ia cancer at diagnosis [15], and a second study also observed no effect on survival among patients with Stage III and IV non-small cell lung cancer at diagnosis.[16] This may indicate physiological differences in the impact of vitamin D on tumor progression at different points in the disease course.

Because of the vitamin D complexes' ability to inhibit proliferation, angiogenesis, and differentiation, Vitamin D could play a promising role in slowing the progression and extending the survival time of esophageal adenocarcinoma patients. To date, the vitamin D pathway has not been studied in relation to EAC survival, so we are not clear on which elements of the pathway are most relevant to prognosis among patients, or subsets of patients, with EAC. This dissertation will explore the role of vitamin D on esophageal cancer prognosis through serum levels that reflecting dietary intake and sun exposure, through potential effect modifiers of the relationship

between 25(OH)D levels and prognosis as well as through mutations in the vitamin D pathway, which are downstream of serum 25(OH)D levels and reflect effect of the vitamin D pathway on tumor progression independent of vitamin D intake. This multifaceted approach allows us to study vitamin D at multiple stages in the pathway to better study its possible, and likely complex, role in EA prognosis.

This dissertation studied the paradoxical relationship between BMI and EAC survival. Aside from being a well-known risk factor for EAC [17-23] and a modifier of the vitamin D pathway[14], obesity has been shown to be associated with longer overall survival in limited studies of EAC.

Due to the relatively small number of cases compared to other types of cancer, survival studies of EAC have been limited to date. In this dissertation, I utilized the ongoing esophageal cancer study from Massachusetts General Hospital, the Molecular Epidemiology of Esophageal Cancer, which contains one of the largest individual study populations of esophageal adenocarcinoma patients, which allows us to test the main effects and also to perform important and relevant subgroup analyses. Given the poor overall survival among esophageal cancer patients, identifiable prognostic markers, and particularly modifiable prognostic factors, are in high demand for this disease. This dissertation delved into a potential role of the modifiable vitamin D pathway in slowing tumor progression as well as attempt to partially elucidate the complex prognostic role of obesity and weight loss prior to diagnosis and their impact on overall survival after diagnosis. The variables studied in this dissertation can serve as prognostic markers and the findings identify future directions for research and clinical treatment of this deadly disease.

Bibliography

1. 2012, G. *Oesophageal Cancer Estimated Incidence, Mortality and Prevalence Worldwide in 2012*. Cancer Fact Sheets 2015 5/10/2017 [cited 2017 10/5/2017]; Available from: http://globocan.iarc.fr/Pages/fact_sheets_cancer.aspx.
2. Zhang, Y., *Epidemiology of esophageal cancer.* World J Gastroenterol, 2013. **19**(34): p. 5598-606.
3. Rubenstein, J.H. and N.J. Shaheen, *Epidemiology, Diagnosis, and Management of Esophageal Adenocarcinoma.* Gastroenterology, 2015. **149**(2): p. 302-17 e1.
4. Blot, W.J., et al., *Rising incidence of adenocarcinoma of the esophagus and gastric cardia.* JAMA, 1991. **265**(10): p. 1287-9.
5. Cook, M.B., W.H. Chow, and S.S. Devesa, *Oesophageal cancer incidence in the United States by race, sex, and histologic type, 1977-2005.* Br J Cancer, 2009. **101**(5): p. 855-9.
6. Holmes, R.S. and T.L. Vaughan, *Epidemiology and pathogenesis of esophageal cancer.* Semin Radiat Oncol, 2007. **17**(1): p. 2-9.
7. Brown, L.M., S.S. Devesa, and W.H. Chow, *Incidence of adenocarcinoma of the esophagus among white Americans by sex, stage, and age.* J Natl Cancer Inst, 2008. **100**(16): p. 1184-7.
8. Pera, M., et al., *Epidemiology of esophageal adenocarcinoma.* J Surg Oncol, 2005. **92**(3): p. 151-9.
9. *Cancer Stat Facts: Esophageal Cancer*. [cited 2017 10/5]; Available from: https://seer.cancer.gov/statfacts/html/esoph.html.
10. El-Serag, H.B. and J. Lagergren, *Alcohol drinking and the risk of Barrett's esophagus and esophageal adenocarcinoma.* Gastroenterology, 2009. **136**(4): p. 1155-7.
11. Anderson, L.A., et al., *The association between alcohol and reflux esophagitis, Barrett's esophagus, and esophageal adenocarcinoma.* Gastroenterology, 2009. **136**(3): p. 799-805.
12. Berlanga-Taylor, A.J. and J.C. Knight, *An integrated approach to defining genetic and environmental determinants for major clinical outcomes involving vitamin D.* Mol Diagn Ther, 2014. **18**(3): p. 261-72.
13. Picotto, G., et al., *Molecular aspects of vitamin D anticancer activity.* Cancer Invest, 2012. **30**(8): p. 604-14.
14. Wortsman, J., et al., *Decreased bioavailability of vitamin D in obesity.* Am J Clin Nutr, 2000. **72**(3): p. 690-3.
15. Zhou, W., et al., *Polymorphisms of vitamin D receptor and survival in early-stage non-small cell lung cancer patients.* Cancer Epidemiol Biomarkers Prev, 2006. **15**(11): p. 2239-45.
16. Heist, R.S., et al., *Circulating 25-hydroxyvitamin D, VDR polymorphisms, and survival in advanced non-small-cell lung cancer.* J Clin Oncol, 2008. **26**(34): p. 5596-602.
17. Grant, W.B., *Lower vitamin-D production from solar ultraviolet-B irradiance may explain some differences in cancer survival rates.* J Natl Med Assoc, 2006. **98**(3): p. 357-64.
18. Hampel, H., N.S. Abraham, and H.B. El-Serag, *Meta-analysis: obesity and the risk for gastroesophageal reflux disease and its complications.* Ann Intern Med, 2005. **143**(3): p. 199-211.

19. Kubo, A. and D.A. Corley, *Body mass index and adenocarcinomas of the esophagus or gastric cardia: a systematic review and meta-analysis.* Cancer Epidemiol Biomarkers Prev, 2006. **15**(5): p. 872-8.
20. Anderson, L.A., et al., *Risk factors for Barrett's oesophagus and oesophageal adenocarcinoma: results from the FINBAR study.* World J Gastroenterol, 2007. **13**(10): p. 1585-94.
21. Abnet, C.C., et al., *A prospective study of BMI and risk of oesophageal and gastric adenocarcinoma.* Eur J Cancer, 2008. **44**(3): p. 465-71.
22. Corley, D.A., A. Kubo, and W. Zhao, *Abdominal obesity and the risk of esophageal and gastric cardia carcinomas.* Cancer Epidemiol Biomarkers Prev, 2008. **17**(2): p. 352-8.
23. Enzinger, P.C. and R.J. Mayer, *Esophageal cancer.* N Engl J Med, 2003. **349**(23): p. 2241-52.

Serum levels of 25-Hydroxyvitamin D at diagnosis are not associated with overall survival in esophageal adenocarcinoma

Elizabeth Loehrer, MSc[1]; Rebecca A. Betensky, PhD[2]; Edward Giovannuci, MD, ScD[3,4]; Li Su[1]; Andrea, Shafer[1]; Bruce Hollis[5], PhD; David Christiani, MD, MPH[1,3]

1. Department of Environmental Health. Harvard T.H. Chan School of Public Health. Boston, MA, United States
2. Department of Biostatistics. Harvard T.H. Chan School of Public Health. Boston, MA, United States
3. Department of Epidemiology. Harvard T.H. Chan School of Public Health. Boston, MA, United States
4. Department of Nutrition. Harvard T.H. Chan School of Public Health. Boston, MA, United States
5. Department of Pediatrics. College of Medicine. Medical University of South Carolina; Charleston, SC, United States

Abstract

Objective: To investigate the prognostic value of serum levels of 25-Hydroxyvitamin D (25(OH)D) at diagnosis for overall survival (OS) time in esophageal adenocarcinoma.

Methods: We utilized serum samples from 463 esophageal adenocarcinoma patients with complete information on relevant confounders and predictors, recruited at the time of their diagnosis at Massachusetts General Hospital between 1999 and 2015. Serum 25(OH)D levels were measured using radioimmunoassay and adjusted for month of blood draw. We used Log rank tests to test the difference in survival curves across quartiles of 25(OH)D level. We used multivariable extended Cox modelling to estimate the hazard ratio (HR) of death by quartile of 25(OH)D, adjusting for age, sex, body mass index (BMI), smoking status, timing of blood draw, year of diagnosis, and treatment, stratifying baseline hazard by clinical stage and modeling surgical resection as a time-dependent covariate. We performed sensitivity analyses to determine if time between diagnosis date and date of blood draw affected these results. We additionally tested for interaction between clinical stage and BMI on the effect of 25(OH)D on overall survival.

Results: The mean 25(OH)D level at diagnosis was 20.7 ng/mL. We found no evidence that OS curves differed across quartiles of 25(OH)D (Log rank p=0.83). In the adjusted extended cox model, we found no evidence the HR for OS among the highest quartile and any other quartile of 25(OH)D differed from 1.0 (Quartile 2 HR 0.95, 95% Confidence Interval (CI) 0.70-1.31; Quartile 3 HR 1.03, 95%CI 0.76-1.39; Quartile 4 HR 1.01, 95%CI 0.73-1.38, global p-value =0.97). Sensitivity analyses demonstrated these results were consistent when accounting for time between diagnosis and blood draw. Moreover, we did not find evidence of interaction between 25(OH)D and clinical stage (p=0.88) or BMI (p=0.43) on OS.

Conclusion: We did not find evidence that serum level of 25(OH)D at time of diagnosis is associated with OS in esophageal adenocarcinoma patients.

1.1 Introduction

Esophageal adenocarcinoma has markedly increased in western countries over the last five decades, particularly in White men, and is now the predominant subtype of esophageal cancer in the U.S. [1-6] Despite some improvements in treatment, esophageal cancer remains a highly deadly disease with an average 5-year survival of less than 20%, accounting for 1.1% of cancer deaths in US women and nearly 4% of cancer deaths among US men.[7, 8] Survival time varies substantially across clinical stages at diagnosis, and clinical stage, together with other stage-related clinical factors like tumor size, is the current best predictor of prognosis in esophageal adenocarcinoma.[1, 7] Nonetheless, even patients diagnosed with stage I disease have a an average 5 year survival less than 50%.[7] Moreover, 80% of cases are diagnosed at later stages (stage II-IV) with few, if any, modifiable factors to improve prognosis after diagnosis.[7] Markers of prognosis from the time of diagnosis, especially factors with the potential to modify the course prognosis, are in high demand both to clarify which patients, including those with early stage, are most likely to have poor clinical outcomes and to help target improvements to current treatment.

In recent years, the vitamin D pathway has gained attention for its robustly demonstrated oncosupressive effects in *in vitro* and *in vivo* studies, including regulating pathways that inhibit proliferation, angiogenesis, and inflammation as well as pathways that promote cell adhesion and induce apoptosis.[9-13] Hypothetically, if biological studies show that a downstream metabolite in the vitamin D pathway is directly involved with regulating oncosuppressive cell signaling, then presumably more intake of vitamin D should generate more downstream regulation and have a protective effect on the development and progression of cancer. With supplements that

are cheap and readily available to the public, vitamin D makes a particularly attractive tool for prognostic intervention, and this, coupled with the biologically plausible mechanisms from experimental studies, has generated widespread interest in understanding the role of the vitamin D pathway in cancer initiation and progression in humans. In practice, 25-hydroxyvitamin D [25(OH)D], an upstream, serum circulating metabolite, is commonly used in epidemiological studies as a marker of vitamin D that is bioavailable for further metabolism because it is stable, consistently measured over time, and reflective of both sun exposure and dietary intake.[10, 14]

In humans, meta-analyses from observational studies and RCTs suggest that higher circulating 25(OH)D at diagnosis and vitamin D supplementation are protective against total cancer-specific mortality [15, 16], and a large Mendelian randomization study of genetic variants in two genes (CYP2R1 and DHCR7) affecting plasma 25(OH)D levels also showed that genetically low 25(OH)D was associated with increased risk of total cancer mortality.[17] However, the results linking 25(OH)D to cancer survival in specific cancer sites and across sites have been inconsistent.[18] Some epidemiological studies of various cancer sites have found that having clinically sufficient levels of 25(OH)D (>30ng/mL) at the time of diagnosis, or even non-deficient 25(OH)D levels (>10ng/mL), is associated with better overall survival and progression free survival.[19-25] The strongest evidence to date linked to colorectal cancer.[18] Serum 25(OH)D levels at diagnosis have been reported to be associated to overall survival in lung, pancreatic, breast, melanoma, and prostate cancer, among others.[22-24, 26-33] One recent study looked at the effect of 25(OH)D levels in esophageal cancer patients, half of whom had adenocarcinoma, and found no association with overall survival, but the blood levels of vitamin D were drawn on average 6 years before cancer diagnosis.[34] The role of circulating 25(OH)D

at time of diagnosis on survival in esophageal adenocarcinoma is not established. There may be true variability in the effect of vitamin D on prognosis across cancer sites due to unique tumor biology at different cancer sites [26, 35, 36], or the effect of vitamin D on survival may be modified by cancer stage, as reported by some epidemiologic studies of other cancer sites, or by other metabolic factors in patients, such as obesity. These potential effect modifiers may be contributing to the modest reproducibility of results of the same cancer site across studies.

To date, studies of the association of serum 25(OH)D on overall survival in esophageal cancer are limited, and the effect of 25(OH)D has not been studied exclusively in relation to esophageal adenocarcinoma survival. In this study, we tested whether higher level of circulating vitamin D (25(OH)D) is associated with longer overall survival time among patients with esophageal adenocarcinoma patients. We additionally examined possible effect modification by tumor stage at diagnosis and body mass index at diagnosis.

1.2. Methods

1.2.1 Study Population

The source population is an existing study, the Molecular Epidemiology of Esophageal Cancer, which consists of esophageal cancer patients who have been recruited since January 1999 from Massachusetts General Hospital (Boston, MA).[37, 38] Patients were >18 years of age with histologically confirmed diagnosis. Written informed consent was obtained from all patients prior to study participation. At the time of enrollment, a trained interviewer conducted an interview with patients to obtain demographic and lifestyle information. Clinical records were used to determine patients' cancer histology, treatment regimen, cancer stage, and performance

status at time of diagnosis. The study population for this analysis was restricted to participants with histologically confirmed esophageal adenocarcinoma who were recruited at the time of their primary diagnosis between 1999 and September 2015 (N=587), which was when serum samples were sent for analysis. For this analysis, we excluded patients who were recruited at the time of cancer recurrence or cancer remission (i.e. their primary treatment was already completed), who had a concurrent cancer, who only presented to MGH for a second opinion, or who were diagnosed with stage 0 disease. Of the eligible patient participants, 495 patients had serum samples available for analysis, and 463 patients with complete information on all confounders who were included in the analyses (Figure 1.1).

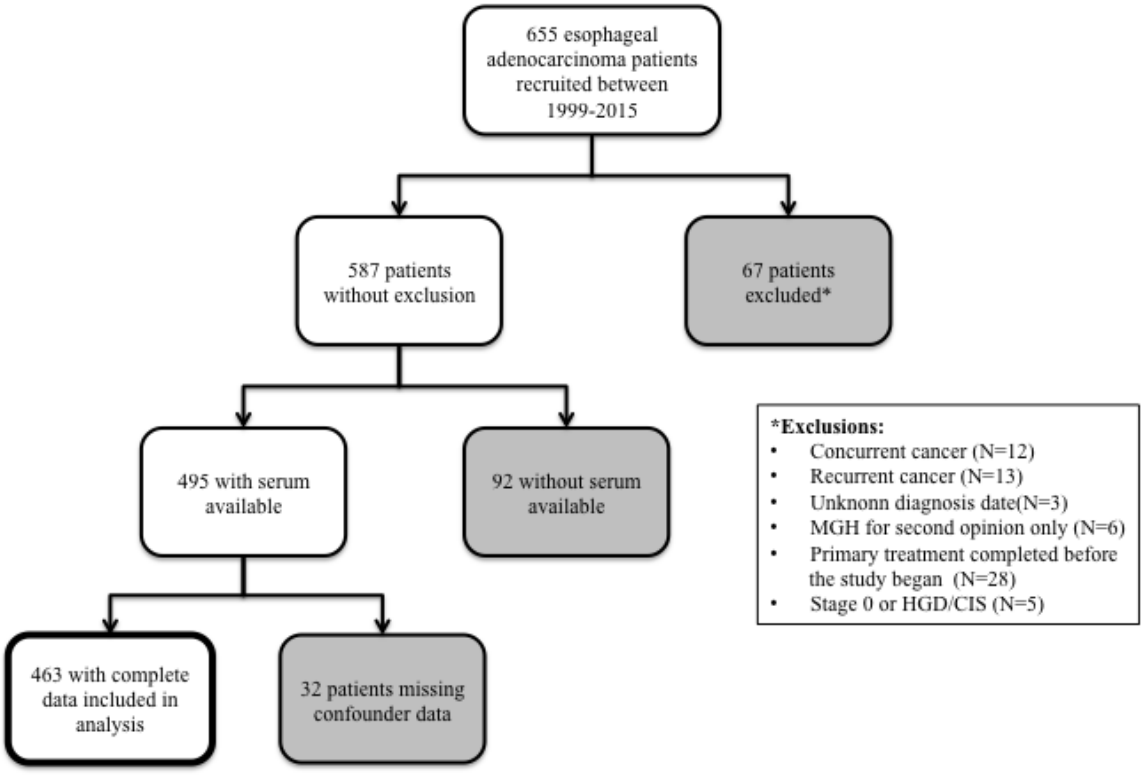

Figure 1.1: Flow chart of the study population

1.2.2 Vitamin D collection and measurement

After patients consented and provided whole blood and serum samples, the samples were stored at 4°C until processing, and were processed within 24 hours of blood draw. Serum was isolated by centrifugation at 2000 r.p.m. for 10min at 4°C. Serum samples were then aliquoted and stored in -80ºC freezers. Between October and November 2015, serum levels of 25(OH)D were measured in the laboratory of Dr. Bruce Hollis (Medical University of South Carolina) by radioimmunoassay (RIA) method.[14, 39] Since serum levels of 25(OH)D are known to fluctuate due to seasonal variability in sun exposure, we generated quartiles of 25(OH)D per month of blood draw, to help reflect seasonal variation. Simulations have shown that this strategy reduces bias toward the null due to measurement misclassification without inducing bias away from the null, which can happen when adjusting for month of blood draw as a covariate in multivariable regression model.[40]

1.2.3 Outcome Variable

The main outcome of interest in this study was overall survival. For the main analysis, overall survival time was defined as the time from date of diagnosis until date of death or censored at date last known to be alive. Data on outcome measures were collected from clinical records and hospital cancer registries. As a sensitivity analysis, we also calculated survival time from the date of blood draw to the date of death or censored at the date last known to be alive, and in this analysis we adjusted for the time between date of diagnosis and the date of blood draw.

1.2.4 Covariate collection and measurement

Information on covariates was collected during the patient's interview and through clinical records. Patients self-reported personal demographic information (age, sex, and race as well as height and weight at diagnosis and smoking history) during the questionnaire. Information regarding cancer stage at diagnosis (defined by TNMG criteria for grouping esophageal adenocarcinoma into clinical staging I-IV and further categorized lymph node negative, lymph node positive, and metastatic), cancer histology, diagnosis date, surgery date (if applicable), treatment regimen, and outcome were obtained from patients' clinical records. Date of diagnosis was considered date of pathology-confirmed cancer. In this study, treatment regimen was modeled as a series of binary variables: chemotherapy (yes/no), radiation (yes/no), and surgery (yes/no), with surgery (i.e. esophagostomy) modeled as a time dependent covariate. Bloods were collected after recruitment into the study. Accounting for the timing of esophagectomy in the model was important for three reasons. First, the timing of the procedure is related to cancer prognosis up to the point of surgery. Patients with early stage tumors will receive esophagectomies as their first treatment whereas patients with advanced metastatic disease will ordinarily not receive the procedure, and locally advanced patients will only receive esophagectomies pending their response to primary chemotherapy and/or radiation treatment. Second, the successful completion of the procedure is the most beneficial form of treatment for esophageal adenocarcinoma patients and one of the best clinical indicators of prognosis. Third, the esophagectomy procedure has a huge impact on patients' diets and weight, and thus likely impacts their vitamin D levels as well. Additionally, we adjusted for year of diagnosis as a continuous variable, to account for possible improvements or slight modifications to treatment protocols throughout the study period. We included crude cigarette smoking history in our

models as an ordinal variable (never, former, current), age was modeled as a continuous variable, and sex was modeled as a dichotomous variable. We used patients' height and weight to calculate BMI at diagnosis (weight(kg) divided by height squared(meters^2)). Once calculated, BMI was categorized into 4 groups: BMI<18.5, 18.5≤BMI<25, 25≤BMI<30, and BMI≥30. Though race, as a proxy for skin pigmentation, is expected to be associated with uptake of Vitamin D through sun exposure and thus associated with circulating Vitamin D levels [41], >90% of our study population identified as White and the remaining participants identified as a variety of races and ethnicities, with each group too small for meaningful statistical comparison across racial categories. Therefore, we ran our main analyses without accounting for race. As a sensitivity analysis, we restricted our study population to patients who identified as White.

1.2.5 Statistical Analyses

1.2.5.1 Main Analyses

We used Kaplan-Meier plots as a univariate visualization of survival time curves between quartiles of serum 25(OH)D and clinical stage. Univariate differences in survival curves were formally tested using Log rank tests. To estimate the association between of serum levels of 25(OH)D and overall survival, we used extended Cox regression models, adjusting for sex, age at diagnosis, BMI, smoking history, year of diagnosis, treatment modality (chemotherapy, radiation, and/or surgery) with surgery modeled as a time-dependent covariate, and stratifying baseline hazard by stage at diagnosis. (Model: $\lambda(t; Z_i(t))=\lambda_{0tumorstage}(t)*\exp(\beta_1*VitD_{q2} + \beta_2*VitD_{q3} + \beta_3*VitD_{q4} + \beta_4*age + \beta_5*sex + \beta_6*DiagnosisYear + \beta_7*Smoking + \beta_8*BMI(1) + \beta_9*BMI(3) + \beta_{10}*BMI(4) + \beta_{11}*Chemotherapy + \beta_{12}*Radiation + \beta_{13}*surgery(t_i))$. When estimating the continuous association of 25(OH)D, we included season of blood draw in the

model. Interactions between serum levels of 25(OH)D quartiles and BMI categories and clinical stage at diagnosis were tested independently by adding interaction terms to the model, in which significance was tested by the joint Wald Test (BMI categories and 25(OH)D quartiles with 8DF, and clinical stage with 25(OH)D quartiles with 6DF).

1.2.5.2 Sensitivity Analyses

Though most of our bloods were taken close to the time of diagnosis, they were not taken exactly at baseline, which means our models included post-baseline measurements of 25(OH)D as if they were baseline measurements, which is problematic in time-to-event analyses. Moreover, many subjects had already initiated treatment at the time of serum draw, and we suspected that chemotherapy, radiation and especially surgical esophagectomy prior to blood draw could have an effect on 25(OH)D levels at the point of time blood is drawn. Thus the timing of blood draw in the year and relative to the diagnosis and any initiation of treatment (for bloods that were drawn longer after diagnosis), 25(OH)D is related to blood draw. Since treatment regimen is decided based on cancer stage and cancer stage and treatment regimen are related to survival, the time of blood draw in relation to treatment regimen is also related to our outcome of overall survival. As a sensitivity analysis to determine that the timing of serum draw was not confounding the estimated effect of serum 25(OH)D on overall survival, we re-ran the previous model, but calculated survival time from the date of blood draw until the date of the outcome, adjusting for the same variables as above and included time between diagnosis and blood draw. We additionally crudely imputed 25(OH)D levels at the time of diagnosis by generating a predictive linear regression model of serum 25(OH)D level as a function of time that included time between diagnosis and blood draw and adjusted for age at blood draw, sex, race, smoking

status, month of blood draw, year of diagnosis, BMI, chemotherapy treatment, radiation treatment, and surgery (if surgery had occurred by the time of blood draw). Using the estimated coefficients from this model, we calculated 25(OH)D level when time between diagnosis and blood draw was equal to 0 (that is when B1=0) for each individual given their covariate profile. This represented their estimated diagnosis level of 25(OH)D. We then modeled the HR of the estimated diagnosis 25(OH)D level on overall survival. Finally, we restricted our study population to patients who identified as non-Hispanic White as a sensitivity analysis of the potential effect of race, a proxy for skin pigmentation, on the effect of 25(OH)D and overall survival. All analyses were performed in SAS 9.4(IBM). P-values were considered significant at an alpha-level of 0.05.

1.3. Results

1.3.1 Study population demographics

Table 1.1 displays the characteristics of our study population, including comparison between patients with and without serum available for analysis. A total of 495 patients had a serum sample available for 25(OH)D level measurement. Median time to serum draw was 7.6 weeks (interquartile range: 2.3-15.7 weeks). Those who did not have serum available for 25(OH)D level measurements were more likely to be female and were less likely to have metastatic disease at diagnosis than the group with serum available. The study population was 88.9% male with a mean age at diagnosis was 63.2 years. The majority of the study population identified as non-Hispanic White, and less than 1.5% of the study population identified as Black, Hispanic, Asian, or Native American. Of the 495 patients with serum samples available, 463 patients had complete information for relevant covariates and were included in the analyses.

Table 1.1: Study population characteristics

	Serum Available (N=495)	Serum Not Available (N=92)
Men	440 (88.9%)	73 (79.4%)
Age	63.2 ± 11.0	64.8 ± 10.9
Race		
White	462 (93.3%)	85 (92.4%)
Black	2 (0.4%)	
Hispanic	6 (1.2%)	2 (2.2%)
Asian	4 (0.8%)	
Native	5 (1.0%)	1 (1.1%)
Ever Smoker	382 (77.2%)	74 (80.4%)
Current Smoker	71 (14.3%)	13 (14.2%)
BMI (kg/m2)	27.4 ± 5.0	27.7 ± 5.7
Stage		
Lymph node negative (I-IIA)	157 (31.7%)	30 (32.6%)
Lymph node positive (IIA-IVA)	240 (48.5%)	48 (52.2%)
Metastatic (IVB)	98 (19.8%)	14 (15.2%)
Treatment		
Surgery alone	87 (17.6%)	22 (23.9%)
Trimodality (chemoradiation and surgery)	261 (52.8%)	42 (45.7%)
Chemoradiation alone	82 (16.6%)	15 (16.3%)
Chemotherapy alone	40 (8.1%)	7 (7.6%)
Radiation therapy alone	7 (1.4%)	3 (3.3%)
Other	17 (3.4%)	3 (3.3%)
25(OH)D (ng/mL)	20.7 ± 10.2	
Death	363 (73.3%)	56 (60.1%)

Values represent number(%) or mean ± SD. Among participants with serum available, information was missing about race (N=16), smoking status (N=10), BMI (N=26), and treatment

(Table 1.1. legend continued) modality (N=2). Among participants with serum not available, information was missing about race (N=4), smoking status (N=1), and BMI (N=3)

1.3.2 Serum 25(OH)D levels and overall survival

The mean 25(OH)D level was 20.7 ng/mL: 11.5% of the study participants had 25(OH)D levels greater or equal to 30ng/mL, and 9.1% had 25(OH)D less than or equal to 10ng/mL. After being categorized into ranked quartiles accounting for month of blood draw, the highest quartile had a mean 25(OH)D of 32.5 ng/mL (SD 11.8), the second quartile had a mean 25(OH)D of 22.2 ng/mL (SD 2.4), the third quartile had a mean 25(OH)D of 17.3 ng/mL (SD 2.8), and the fourth, and lowest, quartile had a mean 25(OH)D of 10.8 (SD 3.8). Figure 1.2 shows a Kaplan-Meier plot to demonstrate the survival curves for each quartile of 25(OH)D levels, and found no difference in overall survival time across them (Figure 1.2; Log rank p=0.83).

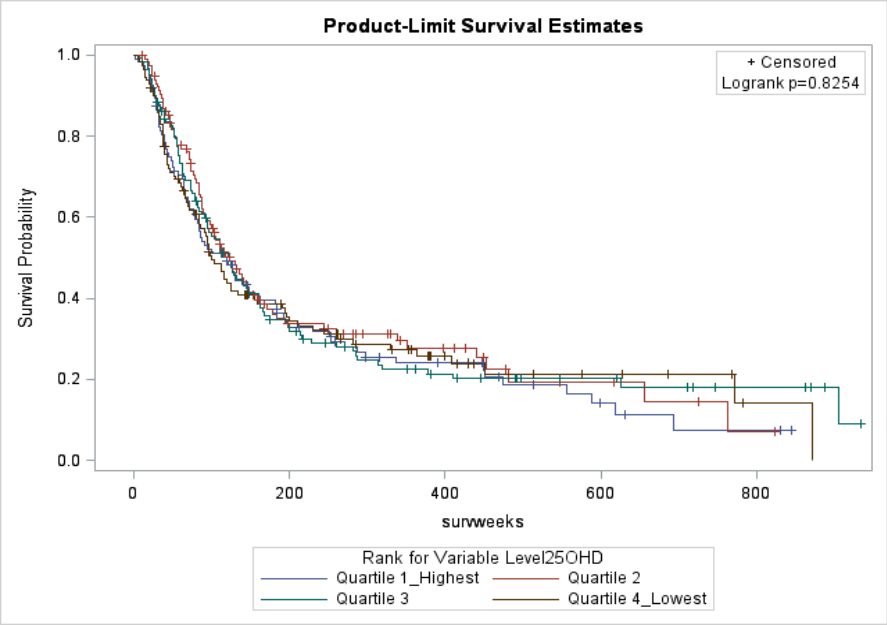

Figure 1.2: Kaplan-Meier survival curve by quartiles of 25(OH)D adjusted for month of blood draw

We used an extended Cox model to estimate the hazard ratio for overall survival by 25(OH)D levels, adjusting for confounders and predictors of survival. When modeled continuously, 25(OH)D showed no association on overall survival (Table 1.2**;** p=0.55). To account for potential nonlinear relationship with overall survival, we modeled quartiles of vitamin D adjusted for month of blood draw, and again found overall that 25(OH)D quartile was not a significant predictor of overall survival (Table 1.2**;** global p=0.97).

Table 1.2: Serum levels of 25(OH)D and overall survival*among EA patients (N=463)

25(OH)D Quartiles†	**Hazard Ratio**	**95% CI**	**p-value**
1 (highest)	REF		
2	0.95	(0.70, 1.31)	0.77
3	1.03	(0.76, 1.39)	0.85
4 (lowest)	1.01	(0.73, 1.38)	0.97
			Global p-value= 0.97
25(OH)D Continuous‡	1.00	(0.99, 1.01)	0.55

*In this analysis, overall survival was calculated as time between date of pathology confirmed diagnosis and date of death or date last known to be alive. †Estimates come from model that additionally adjusted for age, sex, smoking status, BMI categories, year of diagnosis, chemotherapy, radiation, time-dependent surgery, and baseline treatment was stratified by tumor stage by lymph node status. Quartiles of vitamin D were determined accounting for month of blood draw. ‡Estimates come from model that additionally adjusted for age, sex, smoking status, BMI categories, season of blood draw, year of diagnosis, chemotherapy, radiation, time-dependent surgery, and baseline hazard was stratified by tumor stage by lymph node status.

1.3.3 Serum 25(OH)D level and tumor stage interaction with overall survival

Given the strong effect of clinical stage at time of diagnosis on overall survival, we tested the hypothesis that the association between serum 25(OH)D levels on overall survival would be modified by clinical stage. Mean 25(OH)D levels for lymph node negative (19.4 ± 8.1), lymph node positive (20.7 ± 9.4), and metastatic disease (22.4 ± 12.4) at time of diagnosis did not differ significantly (ANOVA p=0.07). Moreover, no significant interaction was found between lymph node status at diagnosis and 25(OH)D levels in the multivariable survival model (Table 1.3; Wald (DF=6), joint test of interaction term, p=0.88).

Table 1.3: Interaction between clinical stage with serum levels of 25(OH)D and overall survival* among EA patients (N=463)

	25(OH)D Quartiles†	HR‡	95% Confidence Limits	
Lymph node negative at diagnosis	Quartile 1 (Highest)	**REF**		
	Quartile 2	0.99	0.51	1.92
	Quartile 3	0.94	0.49	1.79
	Quartile 4 (lowest)	1.03	0.55	1.94
Lymph node positive at diagnosis	Quartile 1 (Highest)	REF		
	Quartile 2	0.98	0.63	1.54
	Quartile 3	1.09	0.71	1.66
	Quartile 4 (lowest)	1.12	0.71	1.76
Metastatic at diagnosis	Quartile 1 (Highest)	REF		
	Quartile 2	0.88	0.50	1.57
	Quartile 3	1.03	0.57	1.86
	Quartile 4 (lowest)	0.81	0.44	1.49

* In this analysis, overall survival was calculated as time between date of pathology confirmed diagnosis and date of death or censored at date last known to be alive. †Quartiles of vitamin D were determined accounting for month of blood draw ‡Model adjusted for the main effect of vitamin D age, sex, smoking status, the main effect of BMI categories, year of diagnosis,

(Table 1.3 legend continued) chemotherapy, radiation, time-dependent surgery, and baseline treatment, and baseline hazard was stratified by tumor stage by lymph node status.

1.3.4 Serum 25(OH)D level and BMI interaction with overall survival

BMI was also considered as a potential modifier of the effect of 25(OH)D on overall survival in esophageal adenocarcinoma patients. Mean 25(OH)D levels for patients with BMI at time of diagnosis <18.5 (15.8 ± 5.9), 18.5≤BMI<25 (20.8 ± 10.1), 25≤BMI<30 (21.6 ± 10.5), and BMI≥30 (19.3 ± 8.5) did not differ significantly (ANOVA p=0.13). We additionally ran a multivariable survival model that included the interaction terms for BMI categories and quartiles of vitamin D adjusted for month of blood draw and found no evidence to support BMI as a modifier of the effect of vitamin D quartile on overall survival (Table 1.4; Wald joint test of interaction term(DF=8), p=0.43).

Table 1.4: Interaction between BMI with serum levels of 25(OH)D and overall survival* among EA patients (N=463)

	25(OH)D Quartiles†	HR‡	95% Confidence Limits	
BMI (<18.5)	Quartile 1 (Highest)	REF		
	Quartile 2	0.30	0.03	3.30
	Quartile 3	0.23	0.02	2.50
	Quartile 4 (lowest)	0.72	0.42	1.23
BMI (≥18.5 and <25)	Quartile 1 (Highest)	REF		
	Quartile 2	0.63	0.36	1.08
	Quartile 3	0.91	0.54	1.53
	Quartile 4 (lowest)	0.72	0.42	1.23
BMI (≥25 and <30)	Quartile 1 (Highest)	REF		
	Quartile 2	1.19	0.75	1.88
	Quartile 3	0.92	0.58	1.45
BMI (≥25 and <30)	Quartile 4 (lowest)	1.25	0.76	2.05
BMI (≥30)	Quartile 1 (Highest)	REF		
	Quartile 2	1.28	0.62	2.65

Table 1.4 Continued				
BMI (≥30)	Quartile 3	1.56	0.78	3.13
	Quartile 4 (lowest)	1.15	0.57	2.32

(Table 1.4 legend continued) *In this analysis, overall survival was calculated as time between date of pathology confirmed diagnosis and date of death or censored at date last known to be alive †Quartiles of vitamin D were determined accounting for month of blood draw. ‡Model estimates adjusted for the main effect of vitamin D age, sex, smoking status, the main effect of BMI categories, year of diagnosis, chemotherapy, radiation, time-dependent surgery, and baseline hazard was stratified by tumor stage by lymph node status.

1.3.5 Sensitivity analysis accounting for time of blood draw

Since the majority of patients did not have their blood drawn at the exact date of diagnosis, we conducted additional sensitivity analyses to determine whether the timing of the blood draw impacted the effect of 25(OH)D level on overall survival in esophageal adenocarcinoma patients. First we tested the effect of quartiles of vitamin D adjusted for month of blood draw on overall survival, taking the date of blood draw as the start of survival time, adjusting for the time between date of diagnosis and date of blood draw. The estimated hazard ratios did not change notably (global p=0.89; Supplementary Table 1.1).

We further estimated patients' 25(OH)D level at the time of diagnosis by creating a linear regression model for 25(OH)D levels at time of blood draw, including a covariate for time between date of diagnosis and blood draw. Because the linear regression coefficient estimates might be unduly influenced by outliers, we further restricted our study population to subjects with 25(OH)D levels with 3 standard deviations (SD) of the mean, and who had their blood

drawn sometime within the week of their diagnosis up to one year past the date of diagnosis. Additionally, given the potentially strong effect of skin pigmentation on vitamin D formation and circulating levels, we further excluded those who were missing information about race. We then used the estimated coefficients from the linear regression model to calculate each participant 25(OH)D level when time between date of diagnosis and blood draw is equal to zero. When we ran the estimated diagnosis date 25(OH)D level in the extended Cox model of overall survival, we again did not see a significant effect of 25(OH)D level on overall survival (Supplementary Table 1.2; p=0.27). We also repeated previous analyses in this restricted study population, and the results did not differ (data not shown).

3.6 Sensitivity analysis restricting to White patients

We additionally repeated all analyses, restricting the population to patients who identified as non-Hispanic White, to determine whether race (as a proxy for skin pigmentation) was not a confounder in our analyses. The results for all analyses did not differ from the models where we used all subjects (Supplementary Table 1.3).

1.4. Discussion

In this study, we did not find evidence that serum levels of 25(OH)D are associated with overall survival in esophageal adenocarcinoma patients. Nor did we find evidence that the effect association of 25(OH)D on overall survival in esophageal adenocarcinoma patients differs by tumor stage at diagnosis, BMI at diagnosis, or timing of the blood draw.

To our knowledge, this is the first study that has assessed the association of diagnostic 25(OH)D levels on overall survival exclusively among esophageal adenocarcinoma patients. A pathologic

study of esophageal tissue samples noted that the vast majority esophageal adenocarcinoma tissues, as well as the precancerous Barrett's esophagus tissues, in the study highly expressed vitamin D receptor (VDR) protein, whereas detection of VDR expression in squamous cell carcinoma was rare, lending evidence that the pathway may be more relevant to the tumor biology in adenocarcinoma than squamous cell carcinoma.[42] At least one study in esophageal squamous cell carcinoma reported patients with >10ng/mL 25(OH)D at time of diagnosis had longer overall survival than those with <10ng/mL. However, they did not clearly report which confounders were controlled for in the models, including if or how they adjusted for the effects of seasonal variability of blood draw.[43] More recently, a study from the European Prospective Investigation in Cancer and Nutrition (EPIC) prospective cohort examined circulating $25(OH)D_3$ levels from blood drawn many years before diagnosis of cancer on mortality among head and neck and esophageal cancer patients.[34] In the 147 esophageal cancer patients included in the analysis (approximately 50% esophageal adenocarcinoma), they found no association between circulating $25(OH)D_3$ levels and overall survival or cancer specific survival.[34] The EPIC study and most studies of 25(OH)D levels and survival time in other cancer sites reported a similar range of 25(OH)D levels to our study population.

There is evidence from human observational studies, Mendelian randomization studies, and RCTs that indicate that low levels of vitamin D are associated with increased risk of total cancer mortality.[15-17] Yet due to the unique biology of different cancers, we do not know whether the vitamin D pathway has the same prognostic impact in all cancer sites. In cancer site specific studies, several studies of colon cancer, breast cancer, and prostate cancer have reported strong associations for poorer survival among patients with very low levels of 25(OH)D (<12ng/mL)

even compared to patients with moderate levels (≥18ng/mL), with one study of colon cancer emphasizing an association in patients with very low levels of 25(OH)D and found no association with overall survival >12ng/mL.[21] In our analyses, patients in the first quartile of 25(OH)D levels would be comparable to these very low levels reported in other cancer sites, but we found no evidence of association between 25(OH)D and overall survival even among patients with extremely low values. Of the studies of specific cancer sites, studies in colorectal cancer arguably provide the strongest and most consistent epidemiologic evidence of the effect of vitamin D on cancer progression and survival.[18] Yet even reports from colorectal cancer are not without nuances. One recent study reported associations between overall survival and free circulating (i.e. not bound to protein in the serum) 25(OH)D and bioavailable 25(OH)D but found no association between total 25(OH)D and overall survival.[20] We only measured total 25(OH)D, we cannot rule out that subsets of 25(OH)D or different molecular markers of the vitamin D pathway may be associated with overall survival or disease specific survival in esophageal adenocarcinoma.

Epidemiologic studies in lung and pancreatic have established nuances to these observed associations by reporting differences in the association of 25(OH)D level on overall survival by stage.[26, 35, 36] Vitamin D is proposed to slow tumor growth and inhibit metastases, among other things, so we hypothesized that the association with 25(OH)D would be weaker among patients with advanced stage disease, once metastases had already occurred. We tested for this effect modification, but we found no evidence of interaction between 25(OH)D and clinical stage on overall survival. Similarly, BMI has a complex relationship with esophageal adenocarcinoma in that higher BMI is a known risk factor developing the disease but a potentially protective

factor for survival after diagnosis.[44-47] BMI is also related to vitamin D, as vitamin D through sun exposure and diet can be stored in the fat cells of a person rather than being converted into the more active 25(OH)D form. Thus people with more adiposity tend to have lower circulating 25(OH)D in their serum than someone with a lower percent body fat but the same level of vitamin D intake.[48] We did not find evidence of BMI as an effect modifier in this study, though we could only use BMI as a measure of body composition. In healthy individuals, BMI is considered a measure of adiposity, but sick cancer patients potentially have underlying sarcopenia or cachexia, so low BMI may indicate low lean mass and not necessarily low adiposity. Thus, the relationship between BMI and vitamin D in esophageal adenocarcinoma cancer survival may be more complex than we were able to capture here.

We acknowledge a number of limitations to our study. First, in this study like many studies, we only have one measure of 25(OH)D from close to the time of diagnosis. Though we accounted for seasonal variability at the time of blood draw, we are still assuming that 25(OH)D levels in patients do not change dramatically over follow-up time, including seasonally, during treatment, or after treatment is completed. Studies in healthy adults have shown that in the absence of taking vitamin D supplementation, 25(OH)D levels do not vary dramatically over a few years, but the levels diverge notably when comparing measurements across a decade, generally decreasing as people age. [49-51] Only a few studies have examined the impact of cancer treatment on levels of 25(OH)D. A recent study of breast cancer reported that patients who received chemotherapy alone had significantly decreased 25(OH)D levels after treatment, but even accounting for that change, neither baseline nor post treatment 25(OH)D levels were associated with pathological response to treatment.[52] In contrast, another recent longitudinal

study of melanoma patients found that baseline levels of 25(OH)D were not associated with risk of relapse but change in vitamin D status during follow-up (both increased and decreased) was associated with worse prognosis, although they did not report change in vitamin D status per specific treatment modality.[53] Our findings cannot rule out that trajectories of 25(OH)D levels throughout treatment may be associated with esophageal adenocarcinoma survival.

A second, related, limitation of our study is that the timing of the blood draw in relation to cancer diagnosis differed across patients, which means although blood draw occurred close to the time of diagnosis, patients were at varying points of their treatment regimen at the time of blood draw. This is a common problem in the study of prognostic biomarkers. We attempted to address this by considering numerous ways in which the timing of the blood draw might have affected measured levels of 25(OH)D, and in all analyses, the timing of blood draw does not appear to have impacted the main results. Third, while we were able to importantly account for the exact timing of esophagectomy (see Section 2.4), we were not able to model chemotherapy and radiation in a time-varying manner due to logistical feasibility limitations. However, patients who receive chemotherapy and radiation will usually receive them before surgery, and thus approximately at baseline. Moreover, the consistently null findings in all analysis and the p-values consistently close to 1 support the null hypothesis, and mean that the potential residual confounding from the above mentioned factors is unlikely to change the result of our analyses.

There are several strengths to our study. To our knowledge, this is the largest study to examine the effect of 25(OH)D levels as a prognostic factor in esophageal adenocarcinoma, and the first to look exclusively at esophageal adenocarcinoma. Our large study population allowed us to

consider possible effect modifiers and many relevant confounders in addition to the main effect. Our patient population was recruited from a large regional cancer center, and our study population demographics are similar to the demographics of esophageal adenocarcinoma patients across the country. Therefore, our findings should be generalizable to the US esophageal adenocarcinoma patients.

Despite the strong biological evidence of the anticancer properties of the vitamin D pathway, the evidence for 25(OH)D serum levels as a marker of prognosis of various cancer sites in humans has been equivocal. Our results do not support that 25(OH)D levels at diagnosis are associated with overall survival in esophageal adenocarcinoma patients. 25(OH)D levels are a more optimal marker of vitamin D status in patients than its downstream, more active metabolite, $1,25(OH)_2D_3$, due to its longer half-life, relative abundance in circulation, and relative stability. However, in terms of relevance for anti-cancer activity, 25(OH)D represents an upstream portion of the vitamin D pathway (namely, a reflection of vitamin D intake and not an indication of how the body is able to metabolize the vitamin D or how the vitamin D regulates cell functions).[14] Other factors, unspecific to the cancer diagnosis itself, such as liver and kidney function, genetic mutations (both germ line and somatic), and genetic expression variability may play a role in altering the metabolism of 25(OH)D to $1,25(OH)_2D_3$ or the level of activity of the vitamin D complex within the tumor cells. Thus, since they were not taken into account in this analysis, we cannot rule out that these downstream factors may mask the role of the vitamin D pathway in esophageal adenocarcinoma tumor progression.

In summary, in this cross-sectional study we did not find evidence that serum vitamin D level is associated with overall survival in esophageal adenocarcinoma. Longitudinal studies tracking the effect of changing vitamin D levels throughout the course of treatment would further inform recommendations for patients. Esophageal cancer is one of the least studied cancers with one of the worst prognoses.[8] Tens of thousands of patients in the US die every year from this disease, and often shortly after their diagnosis.[6] Few prognostic factors are available for patients with esophageal adenocarcinoma, and to date, there is a lack of modifiable factors or specialized treatment for patients with this disease. Efforts should continue to identify the physiologic and molecular processes that might drive different clinical courses for these patients.

Bibliography

1. Rubenstein, J.H. and N.J. Shaheen, *Epidemiology, Diagnosis, and Management of Esophageal Adenocarcinoma.* Gastroenterology, 2015. **149**(2): p. 302-17 e1.
2. Zhang, Y., *Epidemiology of esophageal cancer.* World J Gastroenterol, 2013. **19**(34): p. 5598-606.
3. Blot, W.J., et al., *Rising incidence of adenocarcinoma of the esophagus and gastric cardia.* JAMA, 1991. **265**(10): p. 1287-9.
4. Cook, M.B., W.H. Chow, and S.S. Devesa, *Oesophageal cancer incidence in the United States by race, sex, and histologic type, 1977-2005.* Br J Cancer, 2009. **101**(5): p. 855-9.
5. Brown, L.M., S.S. Devesa, and W.H. Chow, *Incidence of adenocarcinoma of the esophagus among white Americans by sex, stage, and age.* J Natl Cancer Inst, 2008. **100**(16): p. 1184-7.
6. statistics, F.S.A.i.t.f.a.t.S.c., *SEER Survival Data of Relative Survival By Survival Time of Esophageal Cancer between 1988-2013 (SEER 13) for all races (includes Hispanic), Both Sexes, and All Ages.* . Surveilance Research Program, National Cancer Institute.
7. *Cancer Stat Facts: Esophageal Cancer.* [cited 2017 10/5]; Available from: https://seer.cancer.gov/statfacts/html/esoph.html.
8. 2012, G. *Oesophageal Cancer Estimated Incidence, Mortality and Prevalence Worldwide in 2012*. Cancer Fact Sheets 2015 5/10/2017 [cited 2017 10/5/2017]; Available from: http://globocan.iarc.fr/Pages/fact_sheets_cancer.aspx.
9. Berlanga-Taylor, A.J. and J.C. Knight, *An integrated approach to defining genetic and environmental determinants for major clinical outcomes involving vitamin D.* Mol Diagn Ther, 2014. **18**(3): p. 261-72.
10. Picotto, G., et al., *Molecular aspects of vitamin D anticancer activity.* Cancer Invest, 2012. **30**(8): p. 604-14.
11. Engelman, C.D., et al., *Genetic and environmental determinants of 25-hydroxyvitamin D and 1,25-dihydroxyvitamin D levels in Hispanic and African Americans.* J Clin Endocrinol Metab, 2008. **93**(9): p. 3381-8.
12. Wang, Q., et al., *1,25-Dihydroxyvitamin D3 and all-trans-retinoic acid sensitize breast cancer cells to chemotherapy-induced cell death.* Cancer Res, 2000. **60**(7): p. 2040-8.
13. Jiang, F., et al., *Induction of ovarian cancer cell apoptosis by 1,25-dihydroxyvitamin D3 through the down-regulation of telomerase.* J Biol Chem, 2004. **279**(51): p. 53213-21.
14. Holick, M.F., *Vitamin D status: measurement, interpretation, and clinical application.* Ann Epidemiol, 2009. **19**(2): p. 73-8.
15. Keum, N. and E. Giovannucci, *Vitamin D supplements and cancer incidence and mortality: a meta-analysis.* Br J Cancer, 2014. **111**(5): p. 976-80.
16. Chowdhury, R., et al., *Vitamin D and risk of cause specific death: systematic review and meta-analysis of observational cohort and randomised intervention studies.* BMJ, 2014. **348**: p. g1903.
17. Afzal, S., et al., *Genetically low vitamin D concentrations and increased mortality: Mendelian randomisation analysis in three large cohorts.* BMJ, 2014. **349**: p. g6330.
18. Mondul, A.M., et al., *Vitamin D and Cancer Risk and Mortality: State of the Science, Gaps, and Challenges.* Epidemiol Rev, 2017. **39**(1): p. 28-48.
19. Wesa, K.M., et al., *Serum 25-hydroxy vitamin D and survival in advanced colorectal cancer: a retrospective analysis.* Nutr Cancer, 2015. **67**(3): p. 424-30.

20. Yang, L., et al., *Prognostic value of circulating vitamin D binding protein, total, free and bioavailable 25-hydroxy vitamin D in patients with colorectal cancer.* Oncotarget, 2017. **8**(25): p. 40214-40221.
21. Maalmi, H., et al., *Relationship of very low serum 25-hydroxyvitamin D3 levels with long-term survival in a large cohort of colorectal cancer patients from Germany.* Eur J Epidemiol, 2017.
22. Hu, K., et al., *Circulating Vitamin D and Overall Survival in Breast Cancer Patients: A Dose-Response Meta-Analysis of Cohort Studies.* Integr Cancer Ther, 2017: p. 1534735417712007.
23. Yuan, C., et al., *Prediagnostic Plasma 25-Hydroxyvitamin D and Pancreatic Cancer Survival.* J Clin Oncol, 2016. **34**(24): p. 2899-905.
24. Mondul, A.M., et al., *Circulating 25-Hydroxyvitamin D and Prostate Cancer Survival.* Cancer Epidemiol Biomarkers Prev, 2016. **25**(4): p. 665-9.
25. Tretli, S., et al., *Association between serum 25(OH)D and death from prostate cancer.* Br J Cancer, 2009. **100**(3): p. 450-4.
26. Heist, R.S., et al., *Circulating 25-hydroxyvitamin D, VDR polymorphisms, and survival in advanced non-small-cell lung cancer.* J Clin Oncol, 2008. **26**(34): p. 5596-602.
27. McGovern, E.M., et al., *Retrospective analysis of the influence of 25-hydroxyvitamin D on disease progression and survival in pancreatic cancer.* Nutr J, 2016. **15**: p. 17.
28. Lowe, L.C., et al., *Plasma 25-hydroxy vitamin D concentrations, vitamin D receptor genotype and breast cancer risk in a UK Caucasian population.* Eur J Cancer, 2005. **41**(8): p. 1164-9.
29. Kim, Y. and Y. Je, *Vitamin D intake, blood 25(OH)D levels, and breast cancer risk or mortality: a meta-analysis.* Br J Cancer, 2014. **110**(11): p. 2772-84.
30. Webb, P.M., et al., *Circulating 25-hydroxyvitamin D and survival in women with ovarian cancer.* Am J Clin Nutr, 2015. **102**(1): p. 109-14.
31. Bade, B., et al., *Low serum 25-hydroxyvitamin d concentrations are associated with increased risk for melanoma and unfavourable prognosis.* PLoS One, 2014. **9**(12): p. e112863.
32. Shui, I.M., et al., *Circulating vitamin D, vitamin D-related genetic variation, and risk of fatal prostate cancer in the National Cancer Institute Breast and Prostate Cancer Cohort Consortium.* Cancer, 2015. **121**(12): p. 1949-56.
33. Der, T., et al., *Vitamin D and prostate cancer survival in veterans.* Mil Med, 2014. **179**(1): p. 81-4.
34. Fanidi, A., et al., *Circulating vitamin D in relation to cancer incidence and survival of the head and neck and oesophagus in the EPIC cohort.* Sci Rep, 2016. **6**: p. 36017.
35. Zhou, W., et al., *Polymorphisms of vitamin D receptor and survival in early-stage non-small cell lung cancer patients.* Cancer Epidemiol Biomarkers Prev, 2006. **15**(11): p. 2239-45.
36. Cho, M., et al., *Vitamin D deficiency and prognostics among patients with pancreatic adenocarcinoma.* J Transl Med, 2013. **11**: p. 206.
37. Zhai, R., et al., *Whole-miRNome profiling identifies prognostic serum miRNAs in esophageal adenocarcinoma: the influence of Helicobacter pylori infection status.* Carcinogenesis, 2015. **36**(1): p. 87-93.

38. Zhai, R., et al., *Interactions between environmental factors and polymorphisms in angiogenesis pathway genes in esophageal adenocarcinoma risk: a case-only study.* Cancer, 2012. **118**(3): p. 804-11.
39. Hollis, B.W., et al., *Determination of vitamin D status by radioimmunoassay with an 125I-labeled tracer.* Clin Chem, 1993. **39**(3): p. 529-33.
40. Wang, Y., et al., *Comparing methods for accounting for seasonal variability in a biomarker when only a single sample is available: insights from simulations based on serum 25-hydroxyvitamin d.* Am J Epidemiol, 2009. **170**(1): p. 88-94.
41. Gutierrez, O.M., et al., *Racial differences in the relationship between vitamin D, bone mineral density, and parathyroid hormone in the National Health and Nutrition Examination Survey.* Osteoporos Int, 2011. **22**(6): p. 1745-53.
42. Zhou, Z., et al., *Vitamin D receptor is highly expressed in precancerous lesions and esophageal adenocarcinoma with significant sex difference.* Hum Pathol, 2014. **45**(8): p. 1744-51.
43. Gugatschka, M., et al., *Vitamin D status is associated with disease-free survival and overall survival time in patients with squamous cell carcinoma of the upper aerodigestive tract.* Eur Arch Otorhinolaryngol, 2011. **268**(8): p. 1201-1204.
44. Abnet, C.C., et al., *A prospective study of BMI and risk of oesophageal and gastric adenocarcinoma.* Eur J Cancer, 2008. **44**(3): p. 465-71.
45. Yoon, H.H., et al., *Prognostic impact of body mass index stratified by smoking status in patients with esophageal adenocarcinoma.* J Clin Oncol, 2011. **29**(34): p. 4561-7.
46. Kubo, A. and D.A. Corley, *Body mass index and adenocarcinomas of the esophagus or gastric cardia: a systematic review and meta-analysis.* Cancer Epidemiol Biomarkers Prev, 2006. **15**(5): p. 872-8.
47. Lennon, H., et al., *The Obesity Paradox in Cancer: a Review.* Curr Oncol Rep, 2016. **18**(9): p. 56.
48. Wortsman, J., et al., *Decreased bioavailability of vitamin D in obesity.* Am J Clin Nutr, 2000. **72**(3): p. 690-3.
49. Saliba, W., et al., *The longitudinal variability of serum 25(OH)D levels.* Eur J Intern Med, 2012. **23**(4): p. e106-11.
50. Berger, C., et al., *Temporal trends and determinants of longitudinal change in 25-hydroxyvitamin D and parathyroid hormone levels.* J Bone Miner Res, 2012. **27**(6): p. 1381-9.
51. Mirfakhraee, S., et al., *Longitudinal changes in serum 25-hydroxyvitamin D in the Dallas Heart Study.* Clin Endocrinol (Oxf), 2017. **87**(3): p. 242-248.
52. Charehbili, A., et al., *Vitamin D (25-0H D3) status and pathological response to neoadjuvant chemotherapy in stage II/III breast cancer: Data from the NEOZOTAC trial (BOOG 10-01).* Breast, 2016. **25**: p. 69-74.
53. Saiag, P., et al., *Prognostic Value of 25-hydroxyvitamin D3 Levels at Diagnosis and During Follow-up in Melanoma Patients.* J Natl Cancer Inst, 2015. **107**(12): p. djv264.

Genes and SNPs in the Vitamin D pathway and overall survival in esophageal adenocarcinoma

Elizabeth Loehrer[1], Ruyang Zhang[2], Maxine Chen[3], Zhihui Wang[1], Ryan Sun[4], Rebecca A. Betensky[4], Li Su[1], Ed Giovannucci[5,6], David Christiani[1,6]

1. Department of Environmental Health. Harvard T.H. Chan School of Public Health. Boston, MA, United States
2. Department of Biostatistics. Nanjing Medical University School of Public Health. Nanjing, China
3. Department of Medicine. Brigham and Women's Hospital. Harvard T.H. Chan School of Public Health. Boston, MA, United States
4. Department of Biostatistics. Harvard T.H. Chan School of Public Health. Boston, MA, United States
5. Department of Nutrition. Harvard T.H. Chan School of Public Health. Boston, MA, United States
6. Department of Epidemiology. Harvard T.H. Chan School of Public Health. Boston, MA, United States

Abstract

Background: The vitamin D pathway has been implicated in directly and indirectly regulating numerous pathways involved in tumor growth. Single nucleotide polymorphisms (SNPs) in genes that code for proteins involved in the transport and metabolism of vitamin D have been associated with overall survival in several cancers. Esophageal adenocarcinoma (EAC) is a deadly cancer with a shortage of prognostic markers or modifiable prognostic factors. The association between SNPs in the vitamin D pathway and EAC overall survival are unknown.

Methods: We identified 106 tagSNPs among seven candidate genes (*VDR, RXRA, GC, CYP2R1, CYP27A1, CYP27B1, and CYP24A1)* that code for proteins in the vitamin D pathway. We tested the association between individual loci and genes with overall survival among 412 EAC patients from Massachusetts General Hospital who were diagnosed between January 1999 and December 2016. We restricted our analysis to White patients with a primary diagnosis of EAC. We estimated adjusted hazard ratio (HR) of overall survival using an extended Cox model adjusting for age, sex, smoking status, treatment, diagnosis year, and stratifying baseline hazards by stage at diagnosis. We used a set-based test to test the collective association of SNPs from each gene. We corrected for multiple comparisons using FDR, and we considered statistically significant p-FDR<0.05.

Results: There were a total of 335 all-cause deaths during the course of follow-up, with the median time to death 27.4 months (Kaplan-Meier). Three tag SNPs within *CYP24A1* gene (rs2296241 HR=0.83, 95% Confidence Interval(CI) 0.70-0.97; rs927650=0.70, 95% CI 0.70-0.98; rs1570669 HR=1.21, 95%CI 1.02-1.44) and one tag SNP within the *CYP27B1* gene (rs8176345 HR=0.60, 95% CI 0.39-0.94) were marginally statistically significant (unadjusted p<0.05). None of the individual SNPs analyzed were statistically significantly associated with

overall survival in esophageal adenocarcinoma after correcting for multiple testing. In the set-based analysis per gene, *CYP27B1* was marginally significantly associated with overall survival (p=0.052)

Conclusion: While not statistically significant, our results suggest that genetic variants in *CYP24A1* and *CYP27B1*, which code for proteins responsible for regulating the active metabolite of vitamin D, may be associated with overall survival in EAC.

2.1. Introduction

Evidence is mounting that the vitamin D pathway, which is crucial for normal physiological function, plays a critical role in the pathogenesis and progression of several cancers.[1] Within cells in the body, $1,25(OH)_2D_3$, the active metabolite of Vitamin D, forms a complex with several proteins (VDR, RXRA, and RXRB), and this complex directly and indirectly regulates expression of over 200 genes, including genes in numerous pathways that inhibit cancer growth and spread.[2, 3] However, in addition to the proteins that form the $1,25(OH)_2D_3$ complex, numerous proteins are involved the binding and transport (e.g. GC), metabolic activation (e.g. CYP2R1, CYP27A1, CYP27B1) and inactivation (e.g. CYP24A1) of vitamin D after it enters the body through sunlight or dietary intake and before it reaches its active state.[1, 3, 4] Genetic variation within the genes that code for these proteins can alter the form and function of these proteins, and thus, modify the circulating levels of vitamin D and the metabolic efficacy of the vitamin D pathway.[1, 5] Epidemiologic studies have demonstrated that mutations in genes that code for these proteins in the vitamin D pathway are associated with survival in lung, prostate, and breast cancer patients, among others.[6-10] There are some indications that these Vitamin D pathway proteins might be relevant for esophageal adenocarcinoma (EAC) progression as well.

Less than 20% of patients diagnosed with esophageal adenocarcinoma will survive 5 years past their diagnosis.[11, 12] There is a high demand for both modifiable factors that can improve EAC patients' prognosis and biomarkers of progression that can guide targeted approaches to improving treatment. The vitamin D pathway has the potential to provide both for EAC patients. A recent histology-based study found that *VDR* was highly expressed in esophageal adenocarcinoma tumor cells and tissues with precancerous changes.[13] A few epidemiology

studies have also looked at candidate SNPs in the *VDR* gene in relation to esophageal adenocarcinoma risk.[14, 15] However, to date little focus has been placed on the Vitamin D pathway in association to EAC survival, and the role of genes other than *VDR* in the vitamin D pathway have been largely overlooked.

Therefore, we selected 106 tag SNPs in seven candidate genes in the vitamin D pathway (specifically, *VDR, RXRA, GC, CYP2R1, CYP27A1, CYP27B1, and CYP24A1*) based on a literature review of vitamin D genes in cancer survival. We tested whether each individual SNP loci or the gene-based combination of loci variants were associated with overall survival in EAC patients.

2.2. Methods

2.2.1 Study population

The source population is an existing case-control study of esophageal cancer patients who have been recruited since January 1999 from Massachusetts General Hospital (Boston, MA), the Molecular Epidemiology of Esophageal Cancer.[16, 17] Patients were >18 years of age with histologically confirmed diagnosis. Written informed consent was obtained from all patients prior to study participation. At the time of enrollment, a trained interviewer conducted an interview with patients to obtain demographic and lifestyle information. Clinical records were used to determine patients' cancer histology, treatment regimen, cancer stage, and performance status at time of diagnosis. The study population for this analysis was restricted to participants with histologically confirmed esophageal adenocarcinoma who were recruited at the time of their primary diagnosis between 1999 and February 2017 (N=682), when outcome data was last

updated. For this analysis, we excluded patients who were recruited at the time of cancer recurrence or cancer remission (i.e. their primary treatment was already completed), who had a concurrent cancer, who only presented to MGH for a second opinion, or who were diagnosed with stage 0 disease. We additionally restricted our population to patients who identified as White Of the eligible patient participants to avoid confounding by differences allele frequencies by race. 463 participants met the eligible study population definition (Figure 2.1).

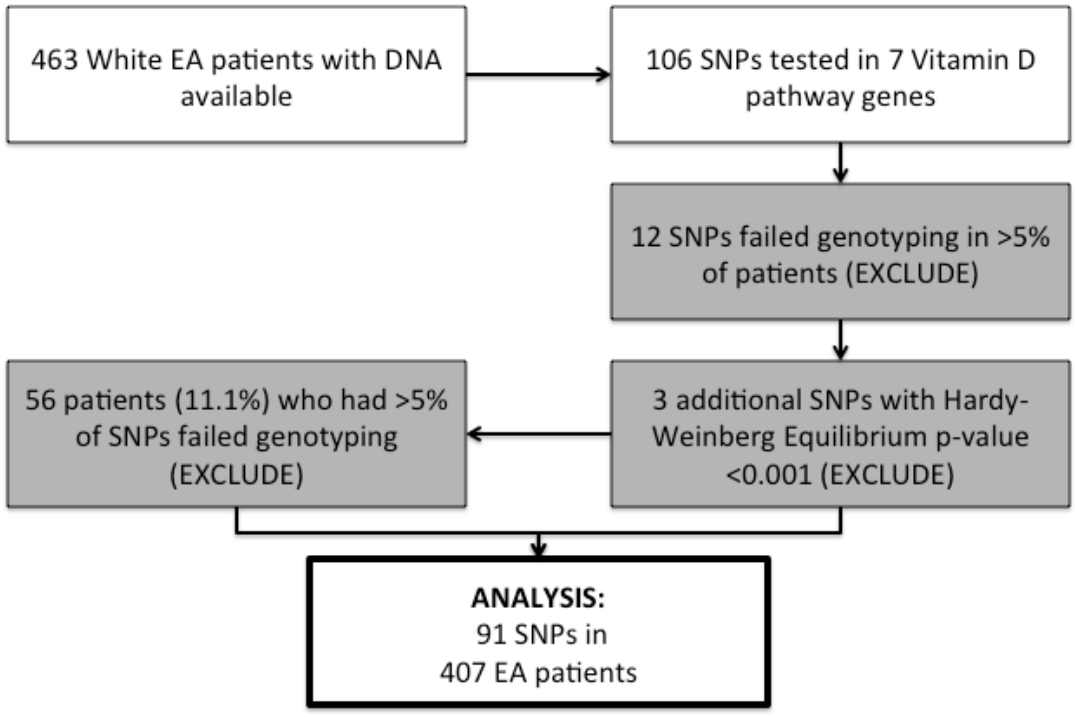

Figure 2.1: Flow chart of study population and quality control steps

2.2.2 Whole blood processing

Patients had a blood and serum tube drawn near the time of recruitment to the study. Samples were stored at 4°C until processing, which was within 24 hours of blood draw. Whole blood was used for DNA extraction.

2.2.3 DNA Extraction

DNA extraction was performed using automated DNA purification through the Autopure LS machine from Qiagen (Qiagen Sciences) on all whole blood samples and on the tissue samples, when available. Whole blood DNA aliquots were stored long term in 4ºC.

2.2.4 Gene and SNP selection

Limited previous work on genetic polymorphisms in the vitamin D pathway has been done in esophageal cancer. Genes were selected based on a literature review of genetic polymorphisms in the vitamin D pathway and cancer survival for other cancer sites. Once the seven genes were selected (*VDR, GC, RXRA CYP24A1, CYP27A1, CYP27B1, CYP2R1*), we used NIEHS LD TAG SNP Selection (TagSNP)[18] to find tag SNPs for each genes with 1000bp flanking the 5' and 3' end of the gene, and including SNPs with a minor allele frequency of >0.05 in Caucasian populations allele frequencies (CEU) and LD threshold of 0.8. A few candidate SNPs in these genes were also included on our list based on findings in previous studies and/or loci with known or predicted functional effects. In the OpenArray design phase, four proposed SNPs were considered invalid candidates for OpenArray and were replaced with other valid SNPs in high LD with the originally proposed SNPs. One additional SNP failed custom assay design and was removed from the list.

2.2.5 Genotyping

Genotyping was provided by the Dana-Farber/Harvard Cancer Center Genotyping and Genetics for Population Sciences core facility, a unit of the Partners HealthCare Center for Personalized Genetic Medicine. Genotyping of the 106 tag SNPs on DNA extracted from whole blood

normalized to 10 ng/uL concentrations was completed using the TaqMan® OpenArray® platform (Life Technologies, Foster City, CA, USA). Samples were plated randomly with respect to time of recruitment and 5% of samples were duplicated for quality control purposes.

2.2.6 Overall survival

Overall survival time was defined as the time from date of pathology-confirmed diagnosis until date of death (all-cause mortality) or date last known to be alive. Data on outcome measure was collected from clinical records and obituary searches.

2.2.7 Covariates

Information on covariates was collected during patient interviews and through clinical records. Personal demographic information age, sex, and race as well as smoking status were self-reported by the patient during the questionnaire. We included crude cigarette smoking history (never, former, current), which was modeled as an ordinal variable in analyses. Information regarding cancer stage at diagnosis (categorized as stage 1-4), cancer histology, diagnosis date, surgery date (if applicable), treatment regimen, and prognosis were obtained from patients' clinical records. Date of diagnosis was considered date of pathology-confirmed cancer. In this study, treatment regimen was modeled as a series of binary variables: chemotherapy (yes/no), radiation (yes/no), and surgery (yes/no). We also adjusted for year of diagnosis to account for possible improvements or slight modifications to treatment protocols throughout the study period. Since the timing of surgery varied between patients and in relation to diagnosis date and

since timing of surgery is related to progression of disease, surgery was modeled as a time-dependent covariate.

2.2.8 Statistical analyses

Kaplan-Meier plots were used to visualize survival curves of patients. Quality control steps of genetic data were conducted in PLINK 1.9. We excluded patients whose genotype call rate was less than 95%, SNPs that failed in more than 5% of participants, and SNPs whose Hardy Weinberg Equilibrium p-value<0.001. During our quality control steps, 12 SNPs failed genotyping in greater than 5% of participants and were excluded (Figure 2.1). An additional 3 SNPs were found to significantly vary from Hardy Weinberg Equilibrium (p<0.001) and were excluded (Figure 1). We also excluded 51 individuals who failed genotyping in greater than 5% of the total SNPs tested (Figure 2.1).

Hazard ratios (HR) for each tag SNPs were estimated using an extended Cox models, adjusting for sex, age at diagnosis, smoking history, year of diagnosis, treatment modality (chemotherapy, radiation, and/or surgery) with surgery modeled as a time-dependent covariate. We also stratified baseline hazard by stage at diagnosis in all models. Gene-based analysis was performed using the Generalized Berk-Jones (GBJ) package in R to test set-based inference.{Sun, 2017 #212} We corrected p-values for multiple testing using the Benjamani-Hochberg False Discovery Rate (FDR) procedure. Statistical analyses were conducted in SAS 9.4 (IBM) and R. Statistical significance for all genetic associations was considered p-FDR<0.05.

2.3. Results

We restricted our study population to patients who identified as White (92.4% of eligible adenocarcinoma patients) to control for confounding by differences in minor allele frequencies of variants by race. Our study population was 88.5% men, 63.4% former smokers, and approximately 50% of the population was lymph node positive at time of diagnosis, with the majority of patients received trimodality as their primary treatment regimen (Table 2.1). There were a total of 335 all-cause mortality events during the course of follow-up. Median follow-up time for the whole population was 27.4 months.

Table 2.1: Patient analysis population characteristics

	Serum Available (N=463)
Men	410 (88.6%)
Age	63.5 ± 11.2
Former Smoker	293 (63.3%)
Current Smoker	64 (13.8%)
BMI (kg/m2)	27.5 ± 5.0
Stage	
Lymph node negative (I-IIA)	151 (32.6%)
Lymph node positive (IIA-IVA)	229 (49.5%)
Metastatic (IVB)	83 (17.9%)
Treatment	
Surgery alone	89 (19.2%)
Trimodality (chemoradiation and surgery)	244 (52.7%)
Chemoradiation alone	72 (15.6%)
Chemotherapy alone	32 (6.9%)

Table 2.1 Continued	
Radiation therapy alone	7 (1.5%)
Other	18 (3.9%)
25(OH)D (ng/mL)	20.8 ± 9.9
All-cause mortality events	335 (72.4%)

Table provides demographic information of White, Adenocarcinoma patients without exclusions, with DNA that was genotyped. Values represent mean ± SD and number (%).

Table 2.2 displays the top results of our individual SNP analyses. Three tag SNPs within *CYP24A1* gene (rs2296241, rs927650, rs1570669) and one tag SNP within the *CYP27B1* gene (rs8176345) were marginally statistically significant (raw p<0.05). The estimated HR for each additional G allele in rs2296241(*CYP24A1)* was 0.83 (95% Confidence Interval(CI) 0.70-0.97, p-FDR=0.68), and each additional T allele of rs927650(*CYP24A1*) or each additional A allele of rs1570669(*CYP24A1*) had a similar estimated HRs. The largest association was observed for rs8176345(*CYP27B1*), where each additional T allele was associated with a 40% reduced hazard of death (Table 2.2 HR=0.60, 95%CI 0.39-0.94, p-FDR=0.68). None of the individual SNPs analyzed were statistically significantly associated with overall survival in esophageal adenocarcinoma after correcting for multiple testing.

Table 2.2: Top SNPs associated with overall survival in esophageal adenocarcinoma patients (Additive Model)

SNP	Minor allele	Gene	HR*	95% CI		Unadjusted p-value	FDR p-value
rs2296241	G	*CYP24A1*	0.83	0.70	0.97	0.02	0.68
rs8176345	T	*CYP27B1*	0.60	0.39	0.94	0.03	0.68
rs927650	T	*CYP24A1*	0.83	0.70	0.98	0.03	0.68
rs1570669	G	*CYP24A1*	1.21	1.02	1.44	0.03	0.68
rs6127119	T	*CYP24A1*	1.21	0.99	1.47	0.06	0.98
rs11185660	C	*RXRA*	0.85	0.70	1.03	0.09	0.98

Table 2.2 Continued							
rs7041	A	*GC*	0.87	0.73	1.03	0.10	0.98
rs4809960	C	*CYP24A1*	1.17	0.96	1.43	0.11	0.98
rs11574032	A	*VDR*	1.27	0.94	1.72	0.11	0.98
rs3755967	T	*GC*	0.87	0.72	1.04	0.13	0.98

*Hazard ratios are adjusted for age, sex, smoking status, diagnosis year, treatment, with surgery modeled as a time dependent co-variate and baseline hazards were stratified by clinical stage

We additionally tested for association between SNPs within a gene and overall survival (Table 3). The set-based test for SNPs in CYP27B1 was marginally associated with overall survival (unadjusted p=0.05), but this was not significant after FDR adjustment. The set-based analysis of SNPs in the other six candidate genes were not significantly associated with overall survival.

Table 2.3:Genes in the vitamin D pathway and overall survival in esophageal adenocarcinoma patients

Gene	Unadjusted p-value	FDR p-value
CYP24A1	0.05	0.37
CYP27B1	0.29	0.99
CYP27A1	0.56	1.00
CYP2R1	0.73	1.00
RXRA	1.00	1.00
GC	1.00	1.00
VDR	1.00	1.00

These are based on the results of set-based test, testing the null hypothesis that there is no association with any SNP in the gene and overall survival among EAC patients.

2.4. Discussion

In a study of candidate genes in the vitamin D pathway, we did not find any individual tag SNPs that were significantly associated with overall survival in esophageal adenocarcinoma patients once adjusting for multiple testing. Four loci in the gene *CYP24A1* and one loci in the gene

CYP27B1 were in the top five most significant results. In the gene-based analysis, *CYP27B1* was marginally associated with overall survival, but we did not observed a significant association with *CYP24A1* or any of the other candidate genes.

The top SNP associations with all-cause mortality in this study were from *CYP24A1* variants rs2296241, rs927650, rs1570669, and rs6127119. CYP24A1 codes for the protein 24-hydroxylase, which is responsible for breaking down 1,25(OH)2D3 into an inactive form as well as breaking down the stored form of 25-hydroxyvitamin D (25(OH)D). 24-hydroxylase is essentially responsible for regulating the amount of active Vitamin D in the body. SNP loci in *CYP24A1* have been reported to associate with risk of developing breast and lung cancer.[19, 20] *CYP24A1* variants have also been associated with increased risk of recurrence and death in prostate cancer.[6] A 2004 study of 42 esophageal tumor (93% squamous cell carcinoma) and adjacent normal tissue found that tumor tissues had relatively increased *CYP24A1* expression compared to normal tissue, and that higher expression of *CYP24A1* was inversely associated with *VDR* expression and overall survival time.[21] A recent study in non-small cell lung cancer found that *CYP24A1* expression was significantly elevated in adenocarcinoma tumors compared to squamous cell carcinoma tumors and in univariate analyses, increased expression of *CYP24A1* and *CYP27B1* was associated with significantly shorter survival time.[22] A similar analyses of *CYP24A1* expression in colorectal tumors found that increased expression was associated with shorter overall survival time and increased risk of recurrence.[23] Given our findings with variants in the *CYP24A1* gene and expression studies of *CYP24A1* in other cancers, we think further study of *CYP24A1* expression and epigenetic regulation in esophageal adenocarcinoma tumor tissues is merited.

CYP27B1 protein is responsible for synthesizing the active form of vitamin D ($1,25(OH)_2D_3$), and thus plays a complimentary role to CYP24A1.[3] Decreased expression of *CYP27B1* in tumor tissues has been shown to be associated with poorer prognosis and shorter survival in melanoma, ovarian, and non-small cell lung cancer.[22, 24, 25] CYP27B1 tagSNP (rs3782130) has also been associated with survival in prostate cancer.[6] We observed a strong association with reduced hazard of death among EAC patients and each additional T allele in tagSNP(rs8176345). rs8176345 has a Regulatory Potential Score of 0.4, indicating a possible role in regulation of *CYP27B1*.[18] These findings in both *CYP24A1* and *CYP27B1* compliment previous studies and indicate that the proteins responsible for regulating the active metabolite of vitamin D may play a role in EAC progression.

Circulating vitamin D levels (25(OH)D) have been suggestively associated with decreased all-cause mortality and cancer specific mortality in various cancer sites.[26-31] Through *in vitro* study of cancer cells, we know that the $1,25(OH)_2D_3$ complex regulation induces cell signaling pathways, promotes cell adhesion to limit proliferation and metastases of cells, and induces apoptosis, and *in vivo* mouse studies have shown that the $1,25(OH)_2D_3$ complex regulation also inhibits angiogenesis and inflammatory effects in tumor cells, all of which have tumor suppressive effects.[3, 32, 33] Increased vitamin D intake does not necessarily correlate with increased levels or activity of the $1,25(OH)_2D_3$ complex within the cell though. Mendelian randomization studies of genes that predict for circulating 25(OH)D levels have been associated with overall survival[5], but epidemiologic studies of 25(OH)D levels and RCTs of vitamin D supplementation have yielded inconsistent findings.[34] These conflicting results indicate that

intake or bioavailable vitamin D may not be a sufficient on their own to improve prognosis. The metabolism and transport of 25(OH)D may be more predictive or essential to see a protective association between circulating 25(OH)D and overall survival in an aggressive tumor like esophageal adenocarcinoma. Our current findings support suggestive associations with variants in multiple genes in the vitamin D pathway, when our previous study found that circulating 25(OH)D was not associated with overall survival in esophageal adenocarcinoma patients (Chapter 2). Targeted inhibition of key proteins in the vitamin D metabolism pathway may prove to have a greater impact on overall survival time, either on its own or combined with vitamin D supplementation for patients.

To our knowledge, this is the first study to examine whether genetic polymorphisms in the vitamin D pathway are associated with overall survival in esophageal adenocarcinoma. Our analyses were bolstered by our thorough clinical information that allowed us to adjust our models for relevant clinical predictors of overall survival. Our study was limited by our modest population size, and we were underpowered to detect low to moderate associations with individual SNPs and correct for multiple comparisons. However, we chose to test a larger number of genes and SNPs because the pathway has not been studied in esophageal adenocarcinoma and it was not clear which genes if any would be most relevant. Therefore we chose to study more genes in the pathway to contribute data to the field. We hope future studies can build on these initial findings.

EAC patients are currently without modifiable factors that could alter their prognosis after diagnosis.[12] Vitamin D supplements are cheap and readily available, making the pathway all

the more alluring for its potential to improve prognosis among cancer patients. However, we are still learning how individual's metabolic differences in combination with differences in tumor biology across cancer sites modifies the association between vitamin D intake and suppressed tumor progression. Different components of the vitamin D pathway may be more or less important in specific cancer sites, and understanding the complex components of the pathway among EAC patients means clinicians can utilize that knowledge of the pathway to best target and improve treatment for EAC patient.

Bibliography

1. Berlanga-Taylor, A.J. and J.C. Knight, *An integrated approach to defining genetic and environmental determinants for major clinical outcomes involving vitamin D.* Mol Diagn Ther, 2014. **18**(3): p. 261-72.
2. Hossein-nezhad, A., A. Spira, and M.F. Holick, *Influence of vitamin D status and vitamin D3 supplementation on genome wide expression of white blood cells: a randomized double-blind clinical trial.* PLoS One, 2013. **8**(3): p. e58725.
3. Picotto, G., et al., *Molecular aspects of vitamin D anticancer activity.* Cancer Invest, 2012. **30**(8): p. 604-14.
4. Du, C., et al., *Pathogenic roles of alterations in vitamin D and vitamin D receptor in gastric tumorigenesis.* Oncotarget, 2017. **8**(17): p. 29474-29486.
5. Afzal, S., et al., *Genetically low vitamin D concentrations and increased mortality: Mendelian randomisation analysis in three large cohorts.* BMJ, 2014. **349**: p. g6330.
6. Holt, S.K., et al., *Vitamin D pathway gene variants and prostate cancer prognosis.* Prostate, 2010. **70**(13): p. 1448-60.
7. Yin, J., et al., *Genetic variants in the vitamin D pathway genes VDBP and RXRA modulate cutaneous melanoma disease-specific survival.* Pigment Cell Melanoma Res, 2016. **29**(2): p. 176-85.
8. Pande, M., et al., *Genetic variants in the vitamin D pathway and breast cancer disease-free survival.* Carcinogenesis, 2013. **34**(3): p. 587-94.
9. Heist, R.S., et al., *Circulating 25-hydroxyvitamin D, VDR polymorphisms, and survival in advanced non-small-cell lung cancer.* J Clin Oncol, 2008. **26**(34): p. 5596-602.
10. Deschasaux, M., et al., *Prospective associations between vitamin D status, vitamin D-related gene polymorphisms, and risk of tobacco-related cancers.* Am J Clin Nutr, 2015. **102**(5): p. 1207-15.
11. *Cancer Stat Facts: Esophageal Cancer*. [cited 2017 10/5]; Available from: https://seer.cancer.gov/statfacts/html/esoph.html.
12. Rubenstein, J.H. and N.J. Shaheen, *Epidemiology, Diagnosis, and Management of Esophageal Adenocarcinoma.* Gastroenterology, 2015. **149**(2): p. 302-17 e1.
13. Zhou, Z., et al., *Vitamin D receptor is highly expressed in precancerous lesions and esophageal adenocarcinoma with significant sex difference.* Hum Pathol, 2014. **45**(8): p. 1744-51.
14. Chang, C.K., et al., *Vitamin d receptor gene variants and esophageal adenocarcinoma risk: a population-based case-control study.* J Gastrointest Cancer, 2012. **43**(3): p. 512-7.
15. Janmaat, V.T., et al., *Vitamin D Receptor Polymorphisms Are Associated with Reduced Esophageal Vitamin D Receptor Expression and Reduced Esophageal Adenocarcinoma Risk.* Mol Med, 2015. **21**: p. 346-54.
16. Zhai, R., et al., *Whole-miRNome profiling identifies prognostic serum miRNAs in esophageal adenocarcinoma: the influence of Helicobacter pylori infection status.* Carcinogenesis, 2015. **36**(1): p. 87-93.
17. Zhai, R., et al., *Interactions between environmental factors and polymorphisms in angiogenesis pathway genes in esophageal adenocarcinoma risk: a case-only study.* Cancer, 2012. **118**(3): p. 804-11.
18. Xu, Z. and J.A. Taylor, *SNPinfo: integrating GWAS and candidate gene information into functional SNP selection for genetic association studies.* Nucleic Acids Res, 2009. **37**(Web Server issue): p. W600-5.

19. Fuhrman, B.J., et al., *Sunlight, polymorphisms of vitamin D-related genes and risk of breast cancer.* Anticancer Res, 2013. **33**(2): p. 543-51.
20. Kong, J., et al., *Genetic polymorphisms in the vitamin D pathway in relation to lung cancer risk and survival.* Oncotarget, 2015. **6**(4): p. 2573-82.
21. Mimori, K., et al., *Clinical significance of the overexpression of the candidate oncogene CYP24 in esophageal cancer.* Ann Oncol, 2004. **15**(2): p. 236-41.
22. Ge, N., et al., *Associations between abnormal vitamin D metabolism pathway function and non-small cell lung cancer.* Oncol Lett, 2017. **14**(6): p. 7538-7544.
23. Sun, H., et al., *CYP24A1 is a potential biomarker for the progression and prognosis of human colorectal cancer.* Hum Pathol, 2016. **50**: p. 101-8.
24. Brozyna, A.A., et al., *Expression of the vitamin D-activating enzyme 1alpha-hydroxylase (CYP27B1) decreases during melanoma progression.* Hum Pathol, 2013. **44**(3): p. 374-87.
25. Brozyna, A.A., et al., *Decreased expression of CYP27B1 correlates with the increased aggressiveness of ovarian carcinomas.* Oncol Rep, 2015. **33**(2): p. 599-606.
26. Wesa, K.M., et al., *Serum 25-hydroxy vitamin D and survival in advanced colorectal cancer: a retrospective analysis.* Nutr Cancer, 2015. **67**(3): p. 424-30.
27. Yang, L., et al., *Prognostic value of circulating vitamin D binding protein, total, free and bioavailable 25-hydroxy vitamin D in patients with colorectal cancer.* Oncotarget, 2017. **8**(25): p. 40214-40221.
28. Maalmi, H., et al., *Serum 25-hydroxyvitamin D levels and survival in colorectal and breast cancer patients: systematic review and meta-analysis of prospective cohort studies.* Eur J Cancer, 2014. **50**(8): p. 1510-21.
29. Hu, K., et al., *Circulating Vitamin D and Overall Survival in Breast Cancer Patients: A Dose-Response Meta-Analysis of Cohort Studies.* Integr Cancer Ther, 2017: p. 1534735417712007.
30. Yuan, C., et al., *Prediagnostic Plasma 25-Hydroxyvitamin D and Pancreatic Cancer Survival.* J Clin Oncol, 2016. **34**(24): p. 2899-905.
31. Mondul, A.M., et al., *Circulating 25-Hydroxyvitamin D and Prostate Cancer Survival.* Cancer Epidemiol Biomarkers Prev, 2016. **25**(4): p. 665-9.
32. Wang, Q., et al., *1,25-Dihydroxyvitamin D3 and all-trans-retinoic acid sensitize breast cancer cells to chemotherapy-induced cell death.* Cancer Res, 2000. **60**(7): p. 2040-8.
33. Jiang, F., et al., *Induction of ovarian cancer cell apoptosis by 1,25-dihydroxyvitamin D3 through the down-regulation of telomerase.* J Biol Chem, 2004. **279**(51): p. 53213-21.
34. Mondul, A.M., et al., *Vitamin D and Cancer Risk and Mortality: State of the Science, Gaps, and Challenges.* Epidemiol Rev, 2017. **39**(1): p. 28-48.

Pre-diagnostic Body Mass Index changes and esophageal adenocarcinoma survival

Elizabeth Loehrer, MSc[1]; Edward Giovannuci, MD, ScD[2,3]; Rebecca A. Betensky, PhD[4]; Andrea Shafer[1]; David C Christiani. MD, MPH[1,3]

1. Department of Environmental Health. Harvard T.H. Chan School of Public Health. Boston, MA, United States
2. Department of Nutrition. Harvard T.H. Chan School of Public Health. Boston, MA, United States
3. Department of Epidemiology. Harvard T.H. Chan School of Public Health. Boston, MA, United States
4. Department of Biostatistics. Harvard T.H. Chan School of Public Health. Boston, MA, United States

Abstract

Background: Though high body mass index (BMI) is a known risk factor for esophageal adenocarcinoma (EA), the effect of BMI on EA overall survival remains unclear.

Methods: The source population is the Esophageal Cancer Study, an ongoing case-control and survivorship study of esophageal cancer patients recruited from Massachusetts General Hospital. In September of 2004, the study introduced a new patient questionnaire that included questions on weight throughout adulthood. For this study, we restricted our analysis to histologically confirmed EA who received their initial diagnosis between 9/1/2004 and 12/31/2015 and completed all relevant questions around BMI (290 patients). Our exposures of interest were BMI categories at diagnosis and average adult BMI (<18.5, 18.5-24.99, 25-29.99, >30 kg/m^2) and ΔBMI (BMI at time of diagnosis-average adult BMI), categorized the differences into ≤-2.71kg/m^2 (substantial weight loss), -2.71< and ≤0 kg/m^2 (stable weight), and >0 kg/m^2 (weight gain). We used an extended Cox proportional hazards model to obtain hazard ratios (HR) of BMI at diagnosis, average adult BMI, and ΔBMI, stratifying baseline hazard by cancer stage, having surgery as a time dependent covariate, and adjusting for radiation and chemotherapy treatment, age, sustained inability to eat solid foods after diagnosis, and smoking status.

Results: There were 186 recorded deaths during follow-up. In this study population, mean age at diagnosis was 63.7 years (SD±9.8), median survival was 2.2 years, mean average adult BMI was 29.0 kg/m^2(SD±5.0), and mean BMI at diagnosis was 27.5 kg/m^2(SD±4.7). Compared to obese subjects at diagnosis, underweight, healthy weight, and overweight at diagnosis respectively had an adjusted hazard ratios of death (HR) of 2.15(95% CI: 0.74-6.25), 1.52(95%, CI: 1.01-2.27) and 0.89 (95% CI: 0.61-1.31) (BMI category p=0.01). Adjusting for confounders, the ΔBMI model shows that patients who lost substantial weight by diagnosis had the highest hazard of

death, compared to patients with stable weight prior to diagnosis (HR=1.77, 95% CI: 1.24-2.53). This association remained after adjusting for weight at diagnosis and weight loss 6 months before diagnosis.

Conclusion: Substantial change in BMI prior to diagnosis indicates poor overall survival in esophageal adenocarcinoma patients. The association is modified by starting BMI, and appears worse for subjects whose average adult BMI was $<27.5 kg/m^2$. Our findings support a biologic association between BMI and overall survival in EA, and not just that our association is driven by reverse causation.

3.1 Introduction

Esophageal adenocarcinoma (EA) is the most common histology of esophageal cancer in the western world and fewer than 20% of patients survive five years past their diagnosis.[1-3] Obesity measured as Body Mass Index (BMI)>30 is an established risk factor for EA[4-9], but many studies have paradoxically shown that higher BMI appears to be associated with longer survival in EA. This "Obesity Paradox", i.e. obesity is a risk factor for developing the cancer but protective for prognosis of the cancer, has been observed in a number of cancers and heavily analyzed in recent years[10, 11], and could be the result of numerous methodological biases[12, 13] or possible protective biological mechanisms.[11, 14, 15]

Methodologically, reverse causation is of particular concern in esophageal cancer because obstruction from tumors and side effects of the treatment dramatically impact patients' ability to sustain their normal diet, often leading to dramatic weight loss and malnutrition, and the weight loss can be pronounced in advanced stages of the disease.[16] Additionally, most esophageal adenocarcinoma patients are middle aged or elderly, and independent of their disease, weight loss among the general population elderly is associated with increased frailty and mortality.[17] Thus, higher BMI at diagnosis may simply indicative of better overall health.

To date, most studies of BMI as a prognostic marker in esophageal cancer have been from surgical oncologists and focused on change in weight after esophagectomy or compared preoperative or pretreatment weight to post treatment weight. The results have been mixed but either indicated no association with BMI on prognosis [18-25] or that patients with higher BMI

tended to have better overall survival.[14, 15, 25-27] Because they rely solely on post-diagnostic measures of BMI, these studies raise concerns about reverse causation.

Few studies have examined early weight prior to disease onset or weight change leading up to diagnosis as a prognostic factor in EA. BMI one year prior to diagnosis has been reported to have no association with overall survival [28, 29], but one study showed that high BMI in early adulthood (age 18-25) was associated with worse overall survival in EA [28]. Additionally, at least two studies report that substantial weight loss (>10% of body weight) leading up to diagnosis has also been associated with poor overall survival in EA. [28, 30]

An estimated 80% of esophageal cancer patients will report some weight loss in the previous six months at the time of diagnosis[31, 32](22), but since high BMI is a strong risk factor for EA, most patients are still overweight or obese at the time of diagnosis even though many report substantial weight loss in the months preceding diagnosis. Considering BMI cross-sectionally or considering weight loss only prior to diagnosis both assume that patients' weight in adulthood prior to diagnosis was static, which is often not the case. Moreover, while weight loss is clinically notable six months prior to diagnosis, subclinical, yet nonetheless substantial, changes to metabolism can occur up to two years before diagnosis.[11] Thus, to study the prognostic association of BMI and weight loss in esophageal cancer, the timing of the measurements and the potential interaction of the measurements are particularly important to avoid reverse causation yet still capture a clinically relevant window.

Therefore, we aimed to partially disentangle the obesity paradox in the association between BMI and change in BMI with overall survival among EA patients. Adult BMI, BMI at diagnosis, and weight trajectory are three unique facets with potentially different associations with overall survival. Thus, we first studied the association between overall survival and BMI at diagnosis (d-BMI) as well as self-reported average adult BMI (a-BMI), representing average weight at least five years before diagnosis to extend the time window between diagnoses and further reduce the potential for reverse causation. Additionally, we studied the change in BMI from average adult weight to time of diagnosis (ΔBMI) on overall survival time in esophageal adenocarcinoma patients, and importantly tested whether this association was independent from patients' BMI at diagnosis and weight loss six months prior to diagnosis. Finally, we tested whether the association with ΔBMI and overall survival in EA was modified by average adult BMI.

3.2 Methods

3.2.1 Study population

The study population comes from an ongoing study of esophageal cancer patients that have been recruited since January 1999 from Massachusetts General Hospital (MGH), the Molecular Epidemiology of Esophageal Cancer [33, 34]. Recruited patients were >18 years of age with histologically confirmed diagnosis. Written informed consent was obtained from all patients prior to study participation. At the time of enrollment, a trained interviewer interviews patients to obtain demographic and lifestyle information through questionnaire. In September 2004, the study introduced a new patient questionnaire that included questions on weight throughout adulthood. The present study was restricted to histologically confirmed EA patients diagnosed between 9/1/2004 and 12/31/2015, who received the updated questionnaire (N=407). We

excluded patients who were recruited at the time of cancer recurrence or cancer remission (i.e. their primary treatment was already completed), who had a concurrent cancer, who only presented to MGH for a second opinion, or who were diagnosed with stage 0 disease. Of the 407 EA patients that met these criteria, 290 (71.3%) had complete information on weight throughout adulthood and were included in analyses (Figure 3.1). Clinical records were used to determine patients' diagnosis date, clinical stage, and treatment regimen.

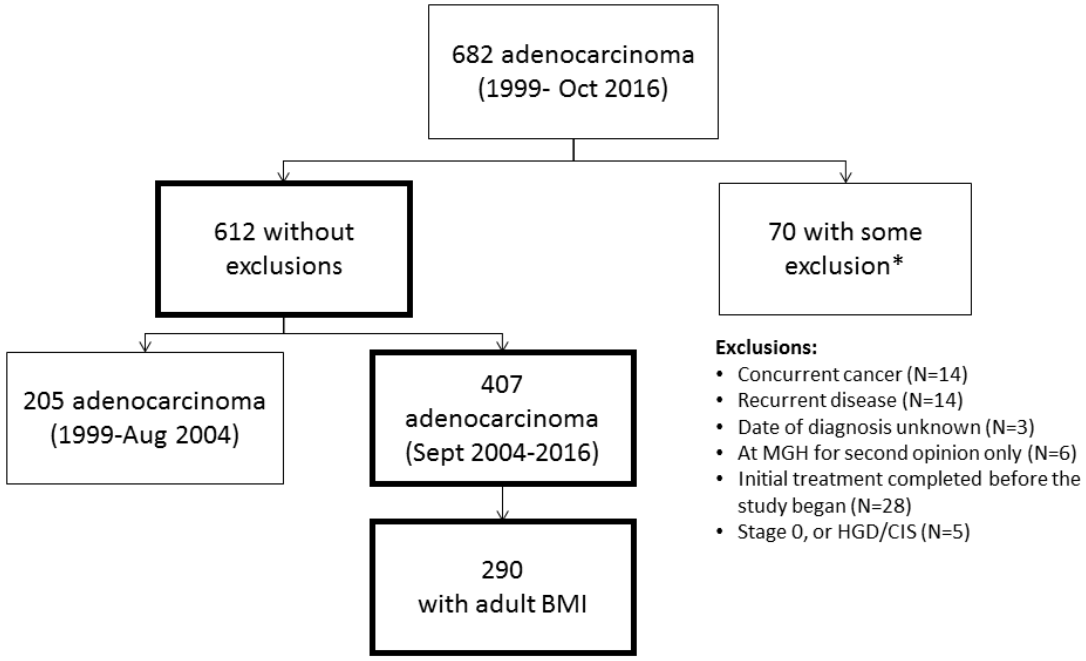

Figure 3.1: Flow chart of the study population

3.2.2 Body mass index measurements

During their questionnaire, patients were asked to report their height and their average weight between age 18-21 years, their average weight between age 21-40 years, their average weight past the age of 40 years, and their weight loss in the 6 months before diagnosis, and their weight

at time of diagnosis. BMI was calculated as weight(kg) divided by height squared(meters^2). Diagnosis BMI (d-BMI) was based on self-reported weight at diagnosis. We considered average adult BMI (a-BMI) to be the patient's self-reported average weight for the time window at least five years before diagnosis. We used this window of time to capture the association of adult BMI on patient prognosis, independent of disease-related weight loss that may occur closer to diagnosis, either directly (e.g. obstruction from tumor, or disease-related cachexia or anorexia) or indirectly (e.g. some patients consciously change their eating and physical activity at the onset of symptoms before pursuing diagnosis). Therefore, patients who were 45 years old or older at time of diagnosis, a-BMI was based on their self-reported average weight past the age of 40 years. For those who were younger than 45 years old at their time of diagnosis, a-BMI was based on their self-reported average weight between 21-40 years. Once calculated, d-BMI and a-BMI were categorized into 4 groups: BMI<18.5 kg/m^2, 18.5≤BMI<25 kg/m^2, 25≤BMI<30 kg/m^2, and BMI≥30 kg/m^2. ΔBMI was defined as the difference between d-BMI and a-BMI. We categorized ΔBMI into tertiles. The 33rd percentile of ΔBMI was -2.71 and the 66th was 0. For easier interpretation, we included 0 in the second category to indicate no change between average adult weight and weight at diagnosis, which we used as our reference. Patients whose ΔBMI≤-2.71 kg/m^2 corresponded to patients who lost substantial weight before diagnosis, relative to their height; Patients with ΔBMI >-2.71 kg/m^2 and ≤0 kg/m^2 corresponded to stable or slight weight loss before diagnosis, relative to height; and Patients with ΔBMI>0 corresponds to weight gain by the time of diagnosis, relative to height.

3.2.3 Overall survival time

Overall survival time was defined as the time from date of pathology-confirmed diagnosis until date of death (all-cause mortality) or censored at date last known to be alive. Data on outcome measure was collected from clinical records and hospital cancer registries.

3.2.4 Covariates

Information on covariates was collected during patient interviews and through clinical records. Personal demographic information age, sex, and race as well as smoking status were self-reported by the patient during the questionnaire. We included crude cigarette smoking history (never, former, current), which was modeled as an ordinal variable in analyses. Information regarding cancer stage at diagnosis (categorized as stage 1-4), cancer histology, diagnosis date, surgery date (if applicable), treatment regimen, and prognosis were obtained from patients' clinical records. Date of diagnosis was considered date of pathology-confirmed cancer. In this study, treatment regimen was modeled as a series of binary variables: chemotherapy (yes/no), radiation (yes/no), and surgery (yes/no). We also adjusted for year of diagnosis to account for possible improvements or slight modifications to treatment protocols throughout the study period. Since the timing of surgery varied between patients and in relation to diagnosis date and since timing of surgery is related to progression of disease, surgery was modeled as a time-dependent covariate.

3.2.5 Statistical models

We used Kaplan-Meier plots to visualize survival time curves between groups. Differences in survival curves were formally tested using Log rank tests. To estimate hazard ratios (HR) of d-

BMI, a-BMI, and ΔBMI on overall survival, we used extended Cox regression models, additionally adjusting for sex, age at diagnosis, smoking history, year of diagnosis, treatment modality (chemotherapy, radiation, and/or surgery) with surgery modeled as a time-dependent covariate. We also stratified baseline hazard by stage at diagnosis in all models. A second model for ΔBMI adjusted for d-BMI in addition to the aforementioned variables. A third model for ΔBMI additionally adjusted for percent change in bodyweight in the 6 months before diagnosis. To check for potential effect modification of the effect of ΔBMI by starting BMI, we performed an analysis of ΔBMI stratified by a-BMI ≥27.5 versus a-BMI<27.5. As a sensitivity analysis, we estimated the HR of d-BMI, restricting to ΔBMI1 (patients who lost substantial weight). All analyses were performed in SAS 9.4(IBM). P-values were considered significant at a two-sided alpha-level of 0.05.

3.3 Results

Of the 407 EA patients diagnosed between 9/1/2004 and 12/31/2015, 290 (71.3%) completed their questionnaire on BMI information and could be included in the analyses. The demographics of our study population are presented in Table 3.1. EA patients' age at diagnosis ranged from 32 to 93 years, and the mean age at diagnosis was 63 years. Most patients were male, White, former smokers, and overweight (Table 3.1). Over the course of follow-up, 186 patients (65.3%) died, and the other 99 patients (34.7%) were censored for overall survival outcome. Median follow-up time for all patients was 25.96 months. Median follow-up time for censored patients was 57.0 months.

Table 3.1: Demographics of the study population

	With BMI available (N=290)	**Without BMI available (N=117)**
Men	259 (89.3%)	96 (82.1%)

Table 3.1 (continued): Demographics of the study population

	With BMI available (N=290)	Without BMI available (N=117)
Age	63.0 ± 10.3	64.2 ± 10.2
Race		
White	278 (95.9%)	92 (78.6%)
Black	1 (0.3%)	
Hispanic	4 (1.4%)	1 (0.9%)
Asian	1 (0.3%)	2 (1.7%)
Native	4 (1.4%)	2 (1.7%)
Other	1 (0.3%)	
Missing	1 (0.3%)	20 (17.1%)
Ever Smoker	213 (73.5%)	88 (75.2%)
Current Smoker	29 (10.0%)	32 (27.4%)
missing		15 (12.8%)
Stage		
1	43 (14.8%)	26 (22.2%)
2	76 (26.2%)	32 (27.4%)
3	98 (33.8%)	48 (41.0%)
4	73 (25.2%)	11 (9.4%)
Treatment*		
Surgery	188 (64.8%)	97 (82.9%)
Radiation	205 (70.7%)	83 (70.9%)
Chemotherapy	239 (82.4%)	87 (74.4%)
missing		1 (0.9%)
Deaths	188 (64.8%)	61 (52.1%)
Diagnosis BMI		
<18 kg/m^2	6 (2.1%)	1 (0.9%)
18.5≤ and<25 kg/m^2	81 (27.9%)	27 (23.1%)
25≤ and<30 kg/m^2	128 (44.1%)	41 (35.0%)
30≥ kg/m^2	75 (25.9%)	31 (26.5%)
missing		17 (14.5%)
Average adult BMI		
<18 kg/m^2	1 (0.3%)	
18.5≤ and<25 kg/m^2	38 (13.1%)	
25≤ and<30 kg/m^2	158 (54.5%)	
30≤ kg/m^2	93 (32.1%)	
ΔBMI		
≤-2.71 kg/m^2	97 (33.4%)	
-2.71< and ≤0 kg/m^2	109 (37.6 %)	
>0 kg/m^2	84 (29.0 %)	

Values presented are mean ± standard deviation or absolute number(population %) *Treatment categories are not mutually exclusive

Mean a-BMI among EA patients was 29.0 kg/m^2 (SD ± 5.0). We found no association between categories of a-BMI and overall survival time by Log-rank (Figure 3.2), nor did we find a difference in adjusted HRs for categories of a-BMI (Table 3.2). The population's mean d-BMI was 27.5 kg/m^2 (SD ± 4.7).

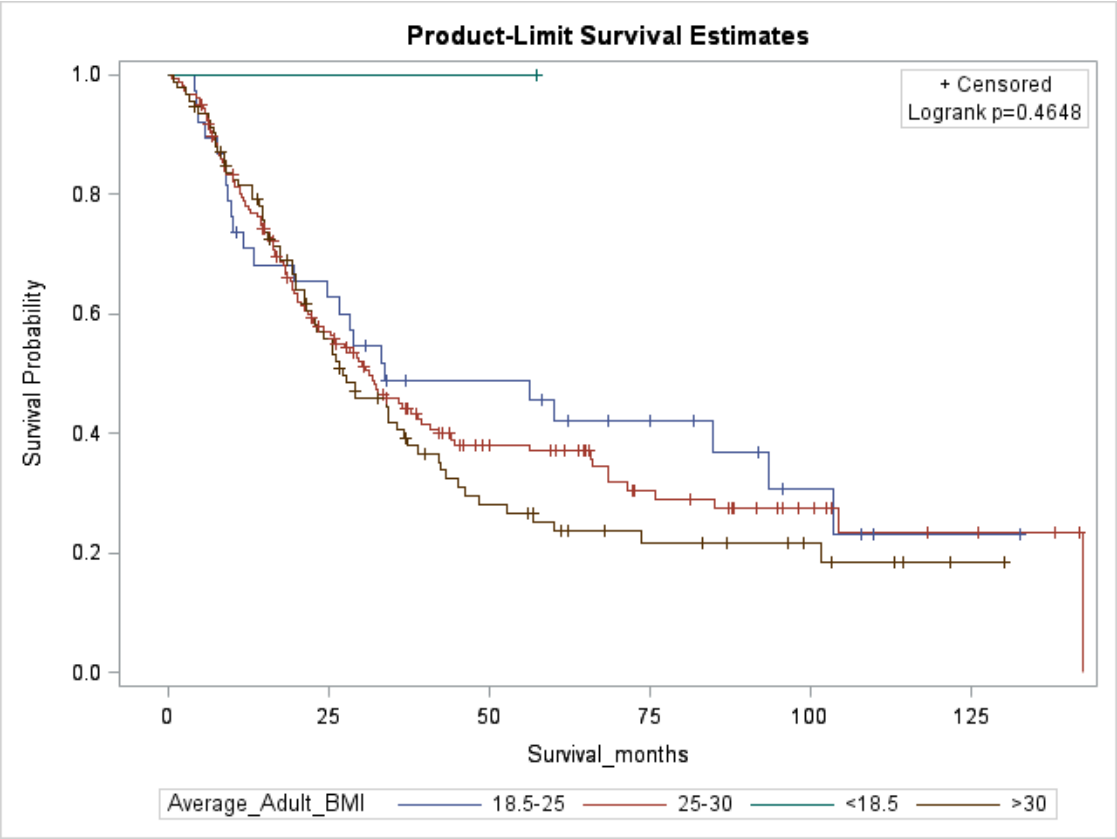

Figure 3.2: Kaplan-Meier plot of survival time by average adult BMI category

Table 3.2: Average adult BMI and overall survival among EA patients

a-BMI categories	Adjusted Hazard Ratio*	95% Confidence Interval	Global p-value
<18 kg/m^2	NA		
18.5≤ and<25 kg/m^2	0.84	(0.51; 1.38)	
25≤ and<30 kg/m^2	0.84	(0.60; 1.17)	
30≤ kg/m^2	REF		P=0.57

(Table 3.2 legend continued)*Model was additionally adjusted for sex, age at diagnosis, smoking status, treatment and year of diagnosis. The model's baseline hazard was stratified by clinical stage at diagnosis, and surgery was coded as a time dependent covariate.

Overall survival time did not differ significantly by categories of d-BMI (Log-rank, p-value=0.15, Figure 3.3).

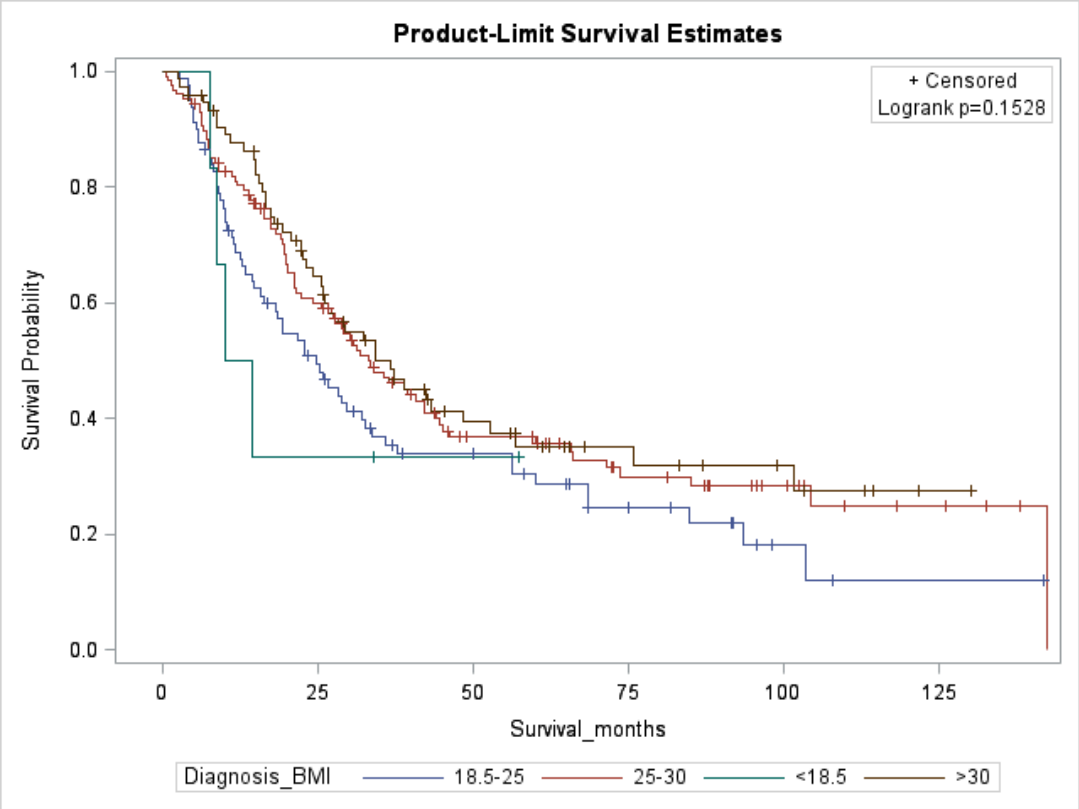

Figure 3.3: Kaplan-Meier plot of survival time by BMI category at time of diagnosis

However, when categories of d-BMI were modeled adjusting for confounders and predictors of the outcome, we saw significant differences between d-BMI categories and hazard of death (Table 3.3). Compared to EA patients with d-BMI ≥30, patients with d-BMI<18.5 kg/m^2 had

higher hazard of death, though not statistically significant (HR: 2.15, 95% confidence interval(CI) (0.74-6.25), p=0.16). Similarly, EA patients d-BMI 18.5≤ and<25 kg/m^2 had an increased hazard of death (HR: 1.52, 95%CI (1.01, 2.27), p=0.04). There was no significant difference in hazard of death between patients with d-BMI 25≤ and<30 kg/m^2 and patients with d-BMI ≥30 (HR: 0.89, 95% CI (0.61, 1.31, p=0.55).

Table 3.3: BMI at time of diagnosis and overall survival among EA patients

d-BMI categories	Adjusted Hazard Ratio*	95% CI	Global p-value
<18 kg/m^2	2.15	(0.74; 6.25)	
18.5≤ and<25 kg/m^2	1.52	(1.01; 2.27)	
25≤ and<30 kg/m^2	0.89	(0.61; 1.31)	
30≤ kg/m^2	REF		P=0.01

*Model was additionally adjusted for sex, age at diagnosis, smoking status, treatment and year of diagnosis. The model's baseline hazard was stratified by clinical stage at diagnosis, and surgery was coded as a time dependent covariate.

In the analysis of ΔBMI, the survival curves differed by category ΔBMI, specifically patients who lost substantial weight relative to their height (ΔBMI≤-2.71 kg/m^2) had poorer overall survival compared to patients with stable weight (-2.71 kg/m^2<ΔBMI≤0 kg/m^2) and patients who gained weight prior to diagnosis (ΔBMI>0 kg/m^2) (Figure 3.4, Log rank p=0.0001).

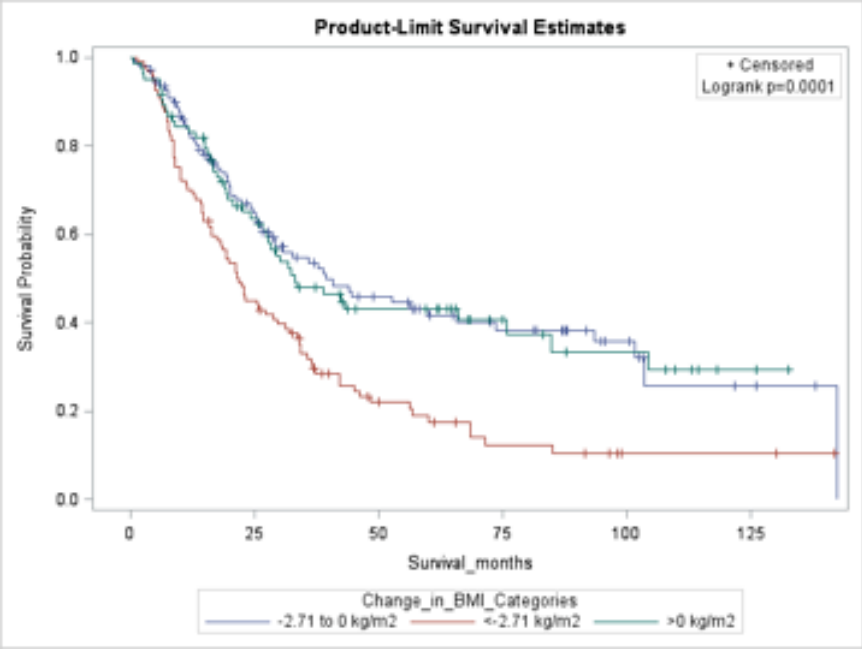

Figure 3.4: Kaplan-Meier plot of survival time by ΔBMI categories

When we adjusted for confounders and predictors of overall survival, both patients who lost substantial weight and patients who gained weight relative to height prior to diagnosis have higher hazards of death compared to patients who had stable BMI (Table 3.4, global p=0.007). The ΔBMI≤-2.71 kg/m^2 group drove the association between ΔBMI and overall survival (HR=1.77; 95% CI 1.24, 2.53; p=0.002). When we additionally adjusted for d-BMI, the association with ΔBMI and overall survival attenuated slightly, but patients who lost substantial weight still had significantly worse hazard of death (Table 3.4, Model 2, HR=1.67; 95% CI 1.15, 2.38; p=0.006). The association between ΔBMI and overall survival did not remain statistically significant when the model was additionally adjusted for percent change in bodyweight in the 6 months before diagnosis, suggesting that the association the association between overall survival and substantial weight loss is driven by weight change leading up to diagnosis, but not entirely explained by it (Table 3.4, Model 3).

Table 3.4: ΔBMI between average adult weight and weight at diagnosis and overall survival among EA patients

ΔBMI Categories (Kg/m^2)	Model 1			Model 2			Model 3		
	Adjusted Hazard Ratio*	95% CI	Global p-value	Adjusted Hazard Ratio†	95% CI	Global p-value	Adjusted Hazard Ratio‡	95% CI	Global p-value
Weight loss: ≤-2.71	1.77	(1.24; 2.53)		1.66	(1.15; 2.38)		1.41	(0.94; 2.11)	
Weight gain: >0	1.27	(0.84; 1.92)		1.29	(0.85; 1.95)		1.29	(0.85; 1.95)	
Stable: >-2.71 and ≤0	Ref		P=0.007	Ref		P=0.02	Ref		0.21

*Model 1: additionally adjusted for sex, age at diagnosis, smoking status, treatment and year of diagnosis. The model's baseline hazard was stratified by clinical stage at diagnosis, and surgery was coded as a time dependent covariate. †Model 2 was additionally adjusted for d-BMI (linear and quadratic term). ‡Model 3 was additionally adjusted for percent change in weight in the 6 months before diagnosis.

Table 3.5 shows the ΔBMI analysis when stratified by a-BMI (<27.5 versus ≥27.5). Among patients who reported a-BMI<27.5, the ΔBMI≤-2.71 kg/m^2 group had an estimated HR of 2.19 (95%CI 1.02, 4.72; p=0.05) compared to patients with stable BMI prior to diagnosis. In contrast, among patients who reported a-BMI ≥27.5, overall survival in ΔBMI≤-2.71 kg/m^2 group did not differ significantly from patients with stable BMI prior to diagnosis (HR=1.21; 95%CI 0.73, 2.02; p=0.46). The hazard of death for patients who gained BMI prior to diagnosis group did not differ significantly from patients with stable BMI in either stratum (Table 3.5).

Table 3.5: ΔBMI between average adult weight and weight at diagnosis and overall survival among EA patients, stratified by overweight or obese as adults

	a-BMI<27.5 (N=120)			**a-BMI≥27.5 (N=170)**		
ΔBMI Categories (Kg/m^2)	N Events/ Patients	Adjusted Hazard Ratio*	95% CI	N Events/ Patients	Adjusted Hazard Ratio*	95% CI
Weight loss: ≤-2.71	25/29	2.19	(1.02; 4.72)	55/68	1.21	(0.73; 2.02)
Weight gain: >0	30/37	1.04	(0.49; 2.20)	31/47	1.15	(0.63; 2.09)
Stable: >-2.71 and ≤0	22/54	Ref		25/55	Ref	
			P=0.13			P=0.75

*Models additionally adjusted for sex, age at diagnosis, smoking status, treatment and year of diagnosis. The baseline hazard was stratified by clinical stage at diagnosis, and surgery was coded as a time dependent covariate.

In a follow-up sensitivity analysis restricted to only patients who lost substantial weight (ΔBMI≤-2.71 kg/m^2), we again found a difference in overall survival time between d-BMI categories, with patients with d-BMI ≥30 kg/m^2 having the lowest hazard of death and patients with d-BMI<18 kg/m^2 having highest hazard of death (Supplementary Table 3.1).

3.4 Discussion

In a survival study of esophageal adenocarcinoma patients, we found that EA patients with BMI<25.0 kg/m2 at the time of diagnosis had an increased hazard rate of all-cause mortality, which is consistent with previous studies that looked only at BMI at the time of diagnosis or post-surgical resection of an esophageal tumor (17-21) Average adult BMI at least five years prior to diagnosis was not associated with overall survival time. This is consistent with previous studies that looked at BMI one year prior to diagnosis (22, 23), and the extended lag time of our adult BMI measure additionally helps control for potential bias that might be induced from subclinical changes related to the tumor.

We are among the first studies to look at the prognostic association of adult weight trajectory and not just disease-associated weight loss prior to diagnosis. We found substantial weight loss between average adult BMI and diagnosis BMI was indicative of poor overall survival compared to patients who maintained stable weight throughout adulthood leading up to diagnosis, independent of BMI at time of diagnosis. We also additionally controlled for weight loss in the 6 months prior to diagnosis, which attenuated the association to non-significant, but did not entirely erase the association. These findings importantly indicate that very sick patients who have low BMI at the time of diagnosis did not entirely explain the association between substantial weight loss and overall survival. In fact, patients with BMI <25kg/m^2 at the time of diagnosis who had stable BMI throughout adulthood had a better prognosis than an overweight or obese patient who lost substantial weight. This is particularly relevant since the majority of people in the substantial weight loss group were still overweight or obese at the time of

diagnosis. Therefore, it is not sufficient to look only at diagnostic BMI and conclude that overweight or high BMI is protective for overall survival.

Most notably, when we stratified ΔBMI by a-BMI, we found that substantial weight loss was associated with significantly worse prognosis in patients with BMI<27.5 kg/m^2 on average as adults than it was for patients with BMI≥27.5 kg/m^2 as adults. While substantial weight loss is bad for both groups, it appears to be worse for patients who had less adiposity and body mass as adults.

These findings may reflect one of two underlying mechanism. If we assume that substantial weight loss in both the adult lean (a-BMI<27.5) and the adult obese (a-BMI≥27.5) represents disease-related weight loss, then our findings may indicate a protective a potential protective effect of extra adiposity for survival. Others have proposed biological hypotheses for the obesity paradox in cancer, including that overweight and obese patients may develop less aggressive tumor subtypes or tumors that are more sensitive to cancer treatment[11] as well as that increased mass or adiposity may provide energy reserves that advantageously help maintain energy and nutrient levels during the harsh treatment for esophageal adenocarcinoma.[10] More likely, the patients who started as lean (a-BMI<27.5) who lost substantial weight leading up to diagnosis were experiencing disease-related weight loss, while patients who started as overweight and obese (a-BMI≥27.5) who lost substantial weight represent a mix of patients who had disease-related weight loss and patients who were intentionally trying to lose weight prior to their diagnosis. If in fact, disease related weight loss is associated with a two-fold risk of all-cause death, then the modest hazard ratio for substantial weight loss among patients with a-

BMI≥27.5 may mean indicate a protective prognostic value of intentional weight loss in adulthood, which drives the harmful association of disease-related weight loss toward the null. However, this analysis cannot clarify what that underlying mechanism might be.

Importantly, we could not take into account body composition in this study. We essentially used BMI as a proxy for adiposity, but BMI is not as correlated with adiposity in the elderly or in cancer patients, who both have tendencies for sarcopenia (low lean muscle mass). Sarcopenia is associated with both cachexia, a cancer-wasting syndrome, and with poor survival in cancer.[32, 35, 36]. Studies of other cancers have indicated that patients with sarcopenic obesity may fair even worse for survival than patients with BMI <25 kg/m^2.[18, 35-38] Thus, accounting for the proportion of adiposity and lean muscle mass in patients is undoubtedly important if we want to assess the driving mechanisms in the association with weight and survival in EA. Studies of prediagnostic body composition and the interplay of body composition and adult weight trajectories in EA are still needed.

Our study has limitations. Although nearly 30% of our population was missing information on adult BMI, we found no difference in overall survival for patients who did and did not report earlier life BMI measures, lowering the concern for selection bias. For our exposure, we relied on self-reported pre-diagnostic average adult weight and diagnostic weight. Diagnosis BMI is expected to be accurate because self-reported current weight has been shown to be accurate [39, 40], and their physicians weigh the patients at time of diagnosis and continually throughout follow-up, so patients will be aware of their diagnosis weight. Studies have shown that self-reported weight from previous times in life is fairly accurate [41], though obese patients may

over-report their average adult weight, in turn underestimating weight gain in adult life.[42] We did not look at BMI continuously, so we expect misclassification on the tails of the patient BMI distribution to have less influence, but misclassification can still occur across BMI categories. In the same note, the older patients were at the time of diagnosis, the more diluted the average adult BMI would be, since the average adult BMI does not capture weight fluctuation or changes in adulthood. However, we had few very old patients at the time of diagnosis and given the aggressive nature of EAC, age at diagnosis is not a strong predictor of overall survival, indicating this misclassification would have a weak impact on our effect estimates.

Our study has numerous strengths in turn. We are among the first to investigate the prognostic value of pre-diagnostic BMI and pre-diagnosis weight change in EAC. We have a relatively large sample size of esophageal adenocarcinoma patients. Rather than use self- reported weight loss in the six months prior to diagnosis, we instead used averaged adult weight because we expect patients will be less likely to associate their average adult weight with their disease to the same extent that they would associate recent weight loss with their disease state Additionally, we have the advantage over many of the previous clinical studies of esophageal survival in that we have thorough demographic and lifestyle data collected systematically for the purposes of research.

Our results highlight the complex interplay of body mass and weight change in esophageal cancer. Given the tendency to focus on cachexia among underweight patients, our findings indicate that more attention is needed to study wasting syndrome among esophageal adenocarcinoma patients who are still clinically categorized as overweight or obese. The role of

body composition and type of mass being loss may help additionally differentiate patient's prognosis and may inform clinical interventions around esophageal adenocarcinoma.

Bibliography

1. Rubenstein, J.H. and N.J. Shaheen, *Epidemiology, Diagnosis, and Management of Esophageal Adenocarcinoma.* Gastroenterology, 2015. **149**(2): p. 302-17 e1.
2. statistics, F.S.A.i.t.f.a.t.S.c., *SEER Survival Data of Relative Survival By Survival Time of Esophageal Cancer between 1988-2013 (SEER 13) for all races (includes Hispanic), Both Sexes, and All Ages.* . Surveilance Research Program, National Cancer Institute.
3. *Cancer Stat Facts: Esophageal Cancer*. [cited 2017 10/5]; Available from: https://seer.cancer.gov/statfacts/html/esoph.html.
4. Zhang, Y., *Epidemiology of esophageal cancer.* World J Gastroenterol, 2013. **19**(34): p. 5598-606.
5. Holmes, R.S. and T.L. Vaughan, *Epidemiology and pathogenesis of esophageal cancer.* Semin Radiat Oncol, 2007. **17**(1): p. 2-9.
6. Cook, M.B., et al., *Childhood body mass index in relation to future risk of oesophageal adenocarcinoma.* Br J Cancer, 2015. **112**(3): p. 601-7.
7. Lagergren, J., R. Bergstrom, and O. Nyren, *Association between body mass and adenocarcinoma of the esophagus and gastric cardia.* Ann Intern Med, 1999. **130**(11): p. 883-90.
8. Thrift, A.P., et al., *Obesity and risk of esophageal adenocarcinoma and Barrett's esophagus: a Mendelian randomization study.* J Natl Cancer Inst, 2014. **106**(11).
9. Smith, M., et al., *Esophageal cancer and body mass index: results from a prospective study of 220,000 men in China and a meta-analysis of published studies.* Int J Cancer, 2008. **122**(7): p. 1604-10.
10. Lennon, H., et al., *The Obesity Paradox in Cancer: a Review.* Curr Oncol Rep, 2016. **18**(9): p. 56.
11. Shachar, S.S. and G.R. Williams, *The Obesity Paradox in Cancer-Moving beyond BMI-Response.* Cancer Epidemiol Biomarkers Prev, 2017. **26**(6): p. 981.
12. Banack, H.R. and J.S. Kaufman, *From bad to worse: collider stratification amplifies confounding bias in the "obesity paradox".* Eur J Epidemiol, 2015. **30**(10): p. 1111-4.
13. Sperrin, M., et al., *Collider Bias Is Only a Partial Explanation for the Obesity Paradox.* Epidemiology, 2016. **27**(4): p. 525-30.
14. Fahey, P.P., et al., *Impact of pre-diagnosis behavior on risk of death from esophageal cancer: a systematic review and meta-analysis.* Cancer Causes Control, 2015. **26**(10): p. 1365-73.
15. Pan, W., et al., *The correlation between high body mass index and survival in patients with esophageal cancer after curative esophagectomy: evidence from retrospective studies.* Asia Pac J Clin Nutr, 2015. **24**(3): p. 480-8.
16. Deans, D.A., et al., *The influence of systemic inflammation, dietary intake and stage of disease on rate of weight loss in patients with gastro-oesophageal cancer.* Br J Cancer, 2009. **100**(1): p. 63-9.
17. Newman, A.B., et al., *Weight change and the conservation of lean mass in old age: the Health, Aging and Body Composition Study.* Am J Clin Nutr, 2005. **82**(4): p. 872-8; quiz 915-6.
18. Martin, L. and P. Lagergren, *Long-term weight change after oesophageal cancer surgery.* Br J Surg, 2009. **96**(11): p. 1308-14.

19. Skipworth, J., et al., *The effect of preoperative weight loss and body mass index on postoperative outcome in patients with esophagogastric carcinoma.* Dis Esophagus, 2009. **22**(7): p. 559-63.
20. Grotenhuis, B.A., et al., *Prognostic value of body mass index on short-term and long-term outcome after resection of esophageal cancer.* World J Surg, 2010. **34**(11): p. 2621-7.
21. Hayashi, Y., et al., *The influence of high body mass index on the prognosis of patients with esophageal cancer after surgery as primary therapy.* Cancer, 2010. **116**(24): p. 5619-27.
22. Blom, R.L., et al., *A high body mass index in esophageal cancer patients does not influence postoperative outcome or long-term survival.* Ann Surg Oncol, 2012. **19**(3): p. 766-71.
23. Shridhar, R., et al., *Body mass index and survival in esophageal adenocarcinoma treated with chemoradiotherapy followed by esophagectomy.* J Gastrointest Surg, 2012. **16**(7): p. 1296-302.
24. Healy, L.A., et al., *Impact of obesity on outcomes in the management of localized adenocarcinoma of the esophagus and esophagogastric junction.* J Thorac Cardiovasc Surg, 2007. **134**(5): p. 1284-91.
25. Scarpa, M., et al., *Overweight patients operated on for cancer of the esophagus survive longer than normal-weight patients.* J Gastrointest Surg, 2013. **17**(2): p. 218-27.
26. Madani, K., et al., *Obesity is not associated with adverse outcome following surgical resection of oesophageal adenocarcinoma.* Eur J Cardiothorac Surg, 2010. **38**(5): p. 604-8.
27. Melis, M., et al., *An elevated body mass index does not reduce survival after esophagectomy for cancer.* Ann Surg Oncol, 2011. **18**(3): p. 824-31.
28. Spreafico, A., et al., *Early adulthood body mass index, cumulative smoking, and esophageal adenocarcinoma survival.* Cancer Epidemiol, 2017. **47**: p. 28-34.
29. Shen, S., et al., *Variation by stage in the effects of prediagnosis weight loss on mortality in a prospective cohort of esophageal cancer patients.* Dis Esophagus, 2017. **30**(9): p. 1-7.
30. van der Schaaf, M.K., et al., *The influence of preoperative weight loss on the postoperative course after esophageal cancer resection.* J Thorac Cardiovasc Surg, 2014. **147**(1): p. 490-5.
31. Laviano, A., et al., *Therapy insight: Cancer anorexia-cachexia syndrome--when all you can eat is yourself.* Nat Clin Pract Oncol, 2005. **2**(3): p. 158-65.
32. Anandavadivelan, P. and P. Lagergren, *Cachexia in patients with oesophageal cancer.* Nat Rev Clin Oncol, 2016. **13**(3): p. 185-98.
33. Zhai, R., et al., *Interactions between environmental factors and polymorphisms in angiogenesis pathway genes in esophageal adenocarcinoma risk: a case-only study.* Cancer, 2012. **118**(3): p. 804-11.
34. Zhai, R., et al., *Whole-miRNome profiling identifies prognostic serum miRNAs in esophageal adenocarcinoma: the influence of Helicobacter pylori infection status.* Carcinogenesis, 2015. **36**(1): p. 87-93.
35. Prado, C.M., et al., *Sarcopenia and cachexia in the era of obesity: clinical and nutritional impact.* Proc Nutr Soc, 2016. **75**(2): p. 188-98.

36. Carneiro, I.P., V.C. Mazurak, and C.M. Prado, *Clinical Implications of Sarcopenic Obesity in Cancer.* Curr Oncol Rep, 2016. **18**(10): p. 62.
37. Gonzalez, M.C., et al., *Obesity paradox in cancer: new insights provided by body composition.* Am J Clin Nutr, 2014. **99**(5): p. 999-1005.
38. Tan, B.H., et al., *Sarcopenia in an overweight or obese patient is an adverse prognostic factor in pancreatic cancer.* Clin Cancer Res, 2009. **15**(22): p. 6973-9.
39. Dahl, A.K., et al., *Agreement between self-reported and measured height, weight and body mass index in old age--a longitudinal study with 20 years of follow-up.* Age Ageing, 2010. **39**(4): p. 445-51.
40. McAdams, M.A., R.M. Van Dam, and F.B. Hu, *Comparison of self-reported and measured BMI as correlates of disease markers in US adults.* Obesity (Silver Spring), 2007. **15**(1): p. 188-96.
41. Troy, L.M., et al., *The validity of recalled weight among younger women.* Int J Obes Relat Metab Disord, 1995. **19**(8): p. 570-2.
42. Norgan, N.G. and N. Cameron, *The accuracy of body weight and height recall in middle-aged men.* Int J Obes Relat Metab Disord, 2000. **24**(12): p. 1695-8.

Conclusion

This dissertation aimed to identify clinical and molecular markers of overall survival among esophageal adenocarcinoma patients who were treated at Massachusetts General Hospital and participated in the Molecular Epidemiology of Esophageal Cancer Study, one of the largest survivor cohort populations of EAC. Specifically, I aimed to study the association between circulating levels of 25(OH)D, mutations among genes coding for proteins in the vitamin D pathway, and how body mass index and change in body mass index before diagnosis are associated with overall survival among EAC patients. Our large study population provided a rich volume of biological and clinical data to study potential molecular markers of EAC prognosis.

Esophageal cancer is one of the least studied cancers.[1] Much of the research to date has focused on risk and detection of the disease. While this is certainly an important strategy for long-term prevention, early detection remains a challenge.[2, 3] In the meantime, tens of thousands of patients die of this disease each year with little hope for survival at the time of diagnosis.[3] Good prognostic biomarkers would help us predict patient's progression free survival and overall survival at their time of diagnosis, identify modifiable factors on which we could intervene to alter a patient's clinical course after diagnosis, and could help drive development of novel treatment strategies for this deadly disease. While the aims of this dissertation would not serve as a cure for the disease, they do have the potential to guide our understanding of the physiologic and molecular differences that might drive different clinical courses in EAC.

For example, if vitamin D levels prove important to prognosis, supplements would be easy to implement into standard treatment and will be worthwhile if it can prolong survival and slow progression. We did not find an association between circulating 25(OH)D levels and overall survival in our population, and the association was not modified by tumor stage or by body mass index, or by timing of the blood draw (Chapter 2). However, we were only able to measure 25(OH)D at the time of diagnosis, and we cannot rule out that trajectory of 25(OH)D levels may be associated with overall survival in EAC. We also did not account for individual metabolic differences in this analysis.

In its active form, $1,25(OH)_2D_3$ forms a complex with VDR, RXRA, and RXRB within the cell, and the complex is responsible for directly and indirectly regulating numerous pathways that suppress tumor promotion and progression, leading to slower tumor growth and longer survival.[4-6] Genetic variants in transport, metabolizing and transcription complex proteins in the vitamin D pathway have been associated with decreased circulating 25(OH)D, increased risk of cancer and poorer prognosis.[5-7] In turn, polymorphisms in vitamin D genes can serve as markers for altered function of the active regulating components of the vitamin D pathway that leads to shorter survival time.[8, 9] These individual metabolic and pathway differences may modify the association between vitamin D intake and circulating 25(OH)D with overall survival.[10] Moreover, tumor biology at different cancer sites might mean that different components of the vitamin D pathway influence tumor progression at different sites.

We were lucky to be among the first to study polymorphisms in the vitamin D pathway in relation to EAC survival. Though none of our results were statistically significant after correction

for multiple comparison (Chapter 3), our top hits occurred among SNPs in the *CYP24A1* gene, which codes for the protein responsible for 24-hydroxylase and 23-hydroxylase of $1,25(OH)_2D$ into the inactive $24,25(OH)_2D$ form, and *CYP27B1* gene, which codes for the protein responsible for 1α-hydroxylase of 25(OH)D into the active form $1,25(OH)_2D$.[4] Both of these proteins are the only established enzymes recognized to carry out these tasks in the vitamin D pathway, and each plays a role in determining the amount of active vitamin D available in the cell [4], making them biologically relevant targets for intervention to modify or enhance the tumor suppressive effects of the vitamin D pathway.

Body mass index can also modify the effects of vitamin D in the body by altering the lipid reservoirs of vitamin D and leading to decreased circulating 25(OH)D for the same amount of intake.[11] Body mass index also appears to play a role in the etiology of EAC as an established risk factor.[12-15] However, the relationship with body mass index and EAC survival has been more complicated to date. Initial studies found that paradoxically, higher BMI is a risk factor for developing EAC but may be protective for survival in EAC.[16-20] This pattern has appeared in other cancers, leading to rigorous association could appear due to methodologically induced bias or due to an underlying biologic mechanism.[21]

My analyses attempted to disentangle some of the complexities of this relationship. I found that indeed just looking at BMI at diagnosis, there was a positive relationship with overall survival, yet substantial weight loss between average adult weight and diagnosis weight (independent of weight at diagnosis) was associated with increased hazard of death (Chapter 4). When we looked only among patients who had lost substantial weight, obese and overweight

patients lower hazard of death and underweight patients had substantially increased hazard of death. We additionally found that the effect of substantial weight loss was modified based on starting weight (average adult BMI), with patients with BMI ≤27.5 kg/m^2 as adults had significantly increased hazard of death with substantial weight loss prior to diagnosis, while patients with BMI >27.5 kg/m^2 had only a slightly increased hazard of death compared to those with stable BMI (Chapter 4). Our findings lend support that a biologic mechanism related to increased body mass plays a protective role for EAC patients. In the future, we hope to further investigate the role of body mass composition in the relationship between body mass and survival in EAC.

Through this dissertation, I was able to explore several molecular and metabolic pathways that seem to be involved in EAC progression and survival. Collectively and individually, these projects provide small, but impactful information to improve our understanding of prognostic factors in this lethal disease. I encourage others to continue the important work of identifying not only targets for improved treatment and prognosis but also essential risk factors that could be targeted to mitigate the incidence of all esophageal cancer in the US and globally.

Bibliography

1. 2012, G. *Oesophageal Cancer Estimated Incidence, Mortality and Prevalence Worldwide in 2012*. Cancer Fact Sheets 2015 5/10/2017 [cited 2017 10/5/2017]; Available from: http://globocan.iarc.fr/Pages/fact_sheets_cancer.aspx.
2. Rubenstein, J.H. and N.J. Shaheen, *Epidemiology, Diagnosis, and Management of Esophageal Adenocarcinoma.* Gastroenterology, 2015. **149**(2): p. 302-17 e1.
3. *Cancer Stat Facts: Esophageal Cancer*. [cited 2017 10/5]; Available from: https://seer.cancer.gov/statfacts/html/esoph.html.
4. Bikle, D.D., *Vitamin D metabolism, mechanism of action, and clinical applications.* Chem Biol, 2014. **21**(3): p. 319-29.
5. Picotto, G., et al., *Molecular aspects of vitamin D anticancer activity.* Cancer Invest, 2012. **30**(8): p. 604-14.
6. Berlanga-Taylor, A.J. and J.C. Knight, *An integrated approach to defining genetic and environmental determinants for major clinical outcomes involving vitamin D.* Mol Diagn Ther, 2014. **18**(3): p. 261-72.
7. Afzal, S., et al., *Genetically low vitamin D concentrations and increased mortality: Mendelian randomisation analysis in three large cohorts.* BMJ, 2014. **349**: p. g6330.
8. Heist, R.S., et al., *Circulating 25-hydroxyvitamin D, VDR polymorphisms, and survival in advanced non-small-cell lung cancer.* J Clin Oncol, 2008. **26**(34): p. 5596-602.
9. Yin, J., et al., *Genetic variants in the vitamin D pathway genes VDBP and RXRA modulate cutaneous melanoma disease-specific survival.* Pigment Cell Melanoma Res, 2016. **29**(2): p. 176-85.
10. Shui, I.M., et al., *Circulating vitamin D, vitamin D-related genetic variation, and risk of fatal prostate cancer in the National Cancer Institute Breast and Prostate Cancer Cohort Consortium.* Cancer, 2015. **121**(12): p. 1949-56.
11. Wortsman, J., et al., *Decreased bioavailability of vitamin D in obesity.* Am J Clin Nutr, 2000. **72**(3): p. 690-3.
12. Hampel, H., N.S. Abraham, and H.B. El-Serag, *Meta-analysis: obesity and the risk for gastroesophageal reflux disease and its complications.* Ann Intern Med, 2005. **143**(3): p. 199-211.
13. Kubo, A. and D.A. Corley, *Body mass index and adenocarcinomas of the esophagus or gastric cardia: a systematic review and meta-analysis.* Cancer Epidemiol Biomarkers Prev, 2006. **15**(5): p. 872-8.
14. Anderson, L.A., et al., *Risk factors for Barrett's oesophagus and oesophageal adenocarcinoma: results from the FINBAR study.* World J Gastroenterol, 2007. **13**(10): p. 1585-94.
15. Abnet, C.C., et al., *A prospective study of BMI and risk of oesophageal and gastric adenocarcinoma.* Eur J Cancer, 2008. **44**(3): p. 465-71.
16. Scarpa, M., et al., *Overweight patients operated on for cancer of the esophagus survive longer than normal-weight patients.* J Gastrointest Surg, 2013. **17**(2): p. 218-27.
17. Madani, K., et al., *Obesity is not associated with adverse outcome following surgical resection of oesophageal adenocarcinoma.* Eur J Cardiothorac Surg, 2010. **38**(5): p. 604-8.
18. Melis, M., et al., *An elevated body mass index does not reduce survival after esophagectomy for cancer.* Ann Surg Oncol, 2011. **18**(3): p. 824-31.

19. Fahey, P.P., et al., *Impact of pre-diagnosis behavior on risk of death from esophageal cancer: a systematic review and meta-analysis.* Cancer Causes Control, 2015. **26**(10): p. 1365-73.
20. Pan, W., et al., *The correlation between high body mass index and survival in patients with esophageal cancer after curative esophagectomy: evidence from retrospective studies.* Asia Pac J Clin Nutr, 2015. **24**(3): p. 480-8.
21. Lennon, H., et al., *The Obesity Paradox in Cancer: a Review.* Curr Oncol Rep, 2016. **18**(9): p. 56.

Supplementary Tables and Figures

Supplementary Table 1.1: Serum levels of 25(OH)D and overall survival (from time of blood draw)* among EA patients (N=463)

25(OH)D Quartiles†	Hazard Ratio	95% CI	p-value
1 (highest)	REF		
2	0.92	(0.67, 1.26)	0.60
3	1.01	(0.74, 1.36)	0.92
4 (lowest)	0.92	(0.67, 1.27)	0.62
			Global p-value= 0.89
25(OH)D Continuous‡	1.00	(0.99, 1.01)	0.63

*In this analysis, overall survival was calculated as time between date of blood draw and date of death or censored at date last known to be alive †Estimates come from model that additionally adjusted for age, sex, smoking status, BMI categories, year of diagnosis, time between diagnosis and blood draw, chemotherapy, radiation, time-dependent surgery, and baseline hazard was stratified by tumor stage by lymph node status. Quartiles of vitamin D were determined accounting for month of blood draw. ‡Estimates come from model that additionally adjusted for age, sex, smoking status, BMI categories, season of blood draw, year of diagnosis, time between diagnosis and blood draw, chemotherapy, radiation, time-dependent surgery, and baseline hazard was stratified by tumor stage by lymph node status

Supplementary Table 1.2: Estimated 25(OH)D at time of diagnosis and overall survival* among EA patients (N=435)

	Hazard Ratio	95% CI	p-value
Estimated 25(OH)D at time of diagnosis continuous†	0.97	(0.92, 1.02)	0.27

* In this analysis, overall survival was calculated as time between date of pathology confirmed diagnosis and date of death or censored at date last known to be alive †Estimates come from model that additionally adjusted for age, sex, smoking status, BMI categories, year of diagnosis, chemotherapy, radiation, time-dependent surgery, and baseline hazard was stratified by tumor stage by lymph node status

Supplementary Table 1.3: Serum levels of 25(OH)D and overall survival* among White EA patients (N=436)

25(OH)D Quartiles†	Hazard Ratio	95% CI	p-value
1 (highest)	REF		
2	0.98	(0.71, 1.35)	0.88
3	0.98	(0.71, 1.34)	0.89
4 (lowest)	0.98	(0.70, 1.35)	0.87
			Global p-value= 0.99
25(OH)D Continuous‡	1.00	(0.99, 1.01)	0.43

*In this analysis, overall survival was calculated as time between date of pathology confirmed diagnosis and date of death or censored at date last known to be alive †Estimates come from model that additionally adjusted for age, sex, smoking status, BMI categories, year of diagnosis, chemotherapy, radiation, time-dependent surgery, and baseline hazard was stratified by tumor stage by lymph node status. Quartiles of vitamin D were determined accounting for month of (Supplementary Table 1.3 Caption continued) blood draw. ‡Estimates come from model that

additionally adjusted for age, sex, smoking status, BMI categories, season of blood draw, year of diagnosis, chemotherapy, radiation, time-dependent surgery, and baseline hazard was stratified by tumor stage by lymph node status

Supplementary Table 2.1: Candidate Gene location, size and number of tag SNPs

Gene	Chromosome	Location (current assembly)	Number of Tag SNPs
VDR	12	47841537..47905031, complement	34
RXRa	9	134326463..134440586	19
GC	4	71741693..71805520, complement	13
CYP24A1	20	54145731..54173985, complement	28
CYP2R1	11	14877436..14898913, complement	7
CYP27A1	2	218781006..218815293	3
CYP27B1	12	57762334..57767193, complement	2

Candidate gene, genomic location and number of tag SNPs used for each gene.

Supplementary Table 3.1: BMI at time of diagnosis and overall survival among only EA patients who lost substantial weight between average adult weight and weight at diagnosis (N=97)

d-BMI categories	Adjusted Hazard Ratio*	95% Confidence Interval	Global p-value
<18 kg/m^2	9.84	(1.68; 57.49)	
18.5≤ and<25 kg/m^2	4.54	(1.36; 15,21)	
25≤ and<30 kg/m^2	2.77	(0.96; 8.03)	
30≤ kg/m^2	REF		P=0.05

*Model was additionally adjusted for sex, age at diagnosis, smoking status, treatment and year of diagnosis. The model's baseline hazard was stratified by clinical stage at diagnosis, and surgery was coded as a time dependent covariate.