VITAMIN A INSUFFICIENCY IN CURRENT AND FORMER SMOKERS AND LUNG CANCER RISK: A CASE-CONTROL STUDY

By

Zachary Bomstein

July, 2021

Director of Thesis: Dr. Richard C. Baybutt, Professor

Major Department: Nutrition Science

ABSTRACT

Background: Adequate dietary intakes of vitamin A are protective against lung cancer, but little is known about the risk of lung cancer when intakes of the vitamin are below minimum physiological requirements. Further investigation into this relationship is needed considering established relationships between smoking, vitamin A deficiency (VAD), and lung cancer risk. *Objective:* The objective of this study was to determine whether intakes of vitamin A below the minimal physiological threshold increased the likelihood of lung cancer. We hypothesized vitamin A intakes above the minimal physiological threshold would reduce the likelihood of lung cancer. *Methods:* A case-control approach was used, utilizing data from eight cycles of NHANES surveys. Cancer-free, everyday cigarette smokers were matched on a 1:1 basis to cases with lung cancer by smoking duration, sex, and age. Paired sample T-tests were used to test differences in mean intakes of retinoids between groups, while an Exact McNemar test was used to assess differences between categorical variables. Conditional logistic regression was used to model the relationship between vitamin A insufficiency and lung cancer incidence, before and after adjustment. *Results*: A covariate in our adjusted model, milk consumption was significantly associated with increased odds of lung cancer (OR = 2.38, CI = 1.04 – 5.43). Contrary to our

hypothesis, intakes of vitamin A at or above the minimal physiological threshold increased the odds of lung cancer, though confounding effects of several covariates, such as the age of smoking initiation, could not be ruled out (OR = 5.49, CI = 0.82 – 36.96). *Conclusions*: While mean intakes of vitamin A and subtypes did not significantly differ between case and control, only 20.4 % of individuals across our sample consumed intakes at or above gender-specific RDAs. As we did not collect data on food-specific micro/macronutrients, more research into dietary contributions of micro/macronutrients from foods known to be implicated in lung cancer prevention or risk, such as vitamin A, vitamin A, calcium, and fatty acids, could prove to be important areas for future research.

VITAMIN A INSUFFICIENCY IN CURRENT AND FORMER SMOKERS AND LUNG CANCER RISK: A CASE-CONTROL STUDY

A Thesis

Presented To the Faculty of the Department of Nutrition Science

East Carolina University

In Partial Fulfillment of the Requirements for the Degree

Master of Science in Nutrition

by

Zachary Bomstein

July, 2021

Vitamin A Insufficiency in Current and Former Smokers and Lung Cancer Risk: A Case-control Study

By

Zachary Bomstein

APPROVED BY:

Director of Thesis: _____

Richard C. Baybutt, PhD

Committee Member _____

Virginia C. Stage, PhD, RDN, LDN

Committee Member _____

Ian Hines, PhD

Committee Member _____

Kevin O'Brien, PhD

Committee Member _____

Name and Degree of Committee Member

Chair of the Department of Nutrition Science _____

Michael D. Wheeler, PhD

Dean of the Graduate School _____

Paul J. Gemperline, PhD

ACKNOWLEDGEMENTS

I would first like to acknowledge Dr. Richard C. Baybutt, who has challenged me as a student and built my confidence as an aspiring researcher in the field of nutrition science. I would next like to thank my committee members, Dr. Virginia C. Stage, Dr. Ian Hines, and Dr. Kevin O'Brien for giving me sound guidance and answering my questions throughout every step of the process. I would further like to thank the committee as a whole for remaining flexible given the uncertainty surrounding the COVD-19 pandemic and the changes made to my research design along the way.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS/ABBREVIATIONS

CHAPTER 1: REVIEW OF THE LITERATURE

1. Vitamin A Sources, Metabolism and Biological Function

Vitamin A is an essential, fat-soluble vitamin responsible for cell/tissue growth and differentiation, vision, epithelial integrity, reproduction, and immune system health [1-3]. In foods, vitamin A is largely found in two forms; as retinyl esters in animal sources and as pro-vitamin A carotenoids in plant sources, with the former known as preformed vitamin A. Collectively, the different forms of vitamin A and its metabolites within the body are known as retinoids [4]. Common sources of preformed vitamin A include liver, egg yolks, and fortified dairy products such as cow's milk. Common sources of provitamin A carotenoids include sweet potatoes, green leafy vegetables, and carrots [4].

Both preformed vitamin A and carotenoids are absorbed into the intestinal epithelium, converted into retinyl esters, and exported into the lymphatic circulation within chylomicrons, which subsequently travel into mainstream circulation. Some of the esters are transferred to specific target tissues and the other retinyl esters remain within the remnants of chylomicrons that are eventually taken up by the liver, where they are stored or hydrolyzed into free retinol [5]. In the liver, vitamin A is stored as retinyl esters where they coalesce to form lipid droplets in storage cells, known as stellate cells. Upon hydrolysis of these stored retinyl esters, retinol is released into the plasma via a transport protein known as retinol-binding protein (RBP); Transthyretin (TTR) attaches to RBP in the plasma, preventing its excretion by the kidneys. The RBP-TTR complex transits the bloodstream where its retinol is taken up by target cells expressing the RBP receptor. Inside target cells, retinol is converted to retinoic acid, the main biologically active retinoid [5-8].

Much of retinoic acid's biological activity can be traced to its interaction with various nuclear receptors, Retinoic Acid Receptors (RAR) and Retinoid X Receptors (RXR). Briefly, upon being shuttled to the nuclear envelope, retinoic acid can interact with these receptors by binding to them, causing conformational changes and, in many cases, dimerization with other receptors; a heterodimeric receptor can then bind to DNA at specific response elements. Next, these response elements act in a specific way to influence basal transcription or chromatin remodeling, actions that are dependent on the coactivator(s) attached to the dimerized receptors [9, 10]. Many of vitamin A's necessary functions can be traced through this genomic activity.

2. Vitamin A Deficiency

2.1. General Deficiency

VAD is a serious threat to human health; its wide range of physiological functions means that in the absence of the vitamin a wide range of physiological processes are disrupted. Collectively, the wide range of symptoms associated with VAD are known as vitamin A Deficiency Disorders or VADD. Along with protein-energy malnutrition, VADD constitutes the world's most common nutritional deficiency [11, 4]. Clinical manifestations of the deficiency include night blindness, dry skin, and abnormal keratinization of mucosal epithelial surfaces, such as those found in the respiratory, genitourinary, gastrointestinal tracts, and cornea [4,12].

In addition to the spectrum of physiological issues VAD causes, maternal VAD appears to have grave consequences on embryonic organ development, likely due to the role of retinoids in regulating cell/tissue differentiation and development; researchers

manipulated doses of dietary vitamin A in pregnant rats to mimic maternal VAD and found that the offspring of vitamin A deficient mothers exhibited severe congenital malformations of the eyes, diaphragm, vasculature, and lungs [13]. In contrast, in humans, in a small sample of Egyptian pregnant women, newborns delivered from mothers with VAD had significantly smaller kidneys compared to those newborns which were delivered from vitamin A sufficient mothers [14]. Additional roles of retinoids in the developing embryo include neural development, epithelial differentiation, craniofacial morphogenesis, and squamous epithelial maintenance [15].

2.2. Deficiency Impacting the Lung

Lung development is a tightly regulated process and retinoic acid (RA) plays an important role in this development. RA coordinates pathways responsible for the genesis of lung budding and regulates alveologenesis largely through RAR activity. Additionally, RA promotes septation, smooth muscle differentiation, and development, and plays an important role in maintaining lung function [16-18]. Consequently, VAD can pose disastrous effects on lung development; in newborn rats, mild VAD induced lower lung weights, lower alveoli septic thickness, and lower total alveoli numbers compared to those which received normal diets [19]. Rats fed a vitamin A-free diet gave birth to rats that exhibited impaired lung growth and development, as well as compromised vital organs such as the heart, liver, and kidney [20]. In populations at risk for VAD and general undernourishment, children with mothers who consumed adequate amounts of vitamin A before, during, and post-pregnancy had better lung function than those children of mothers who consumed a vitamin A-deficient diet during the same period. Subtle defects

in lung development, from VAD or other means, can impact long-term health outcomes of organs and structures of the pulmonary system, like the lungs and alveoli [21, 22].

In addition to causing developmental issues with and within the lung, VAD appears to impact the risk of developing certain conditions affecting the lung as well, such as asthma and emphysema, while increasing the risk of developing certain infections [23]. Vitamin A plays an important role in immune system function and as such, an absence or reduced amount of it may increase the susceptibility to certain infections [23]. In a population of rural Indonesian children, VAD was found to be closely correlated with an increased risk of respiratory disease and diarrhea [24]. In a recent meta-analysis analyzing micronutrient supplementation and risk of various infant infections, vitamin A appeared to exert a protective effect against certain infections, such as measles and bronchopneumonia but failed to show any protective effect against generalized pneumonia [25]. A randomized controlled trial found lower serum vitamin A levels in children with stable asthma compared to healthy children, while a high intake of fruits and antioxidants, as well as adherence to the Mediterranean diet (which is rich in provitamin A carotenoids), was associated with a decreased incidence of asthma [26]. Another study analyzing the relationship between VAD and symptoms of respiratory distress in infants found a positive association between the severity of VAD and the severity of wheezing [27]. Further, airway hyperresponsiveness (AHR) was found to be enhanced in rats fed a vitamin A-deficient diet; the AHR returned to normal after 12 days of RA administration [28].

VAD has also been shown to induce conditions in the lungs of rats similar to that of emphysema with reduced lung elastin content and decreased type II pneumocyte

surfactant synthesis; this is likely due to RA's role in alveolar regeneration and tissue repair [29]. Damage to the oxidant/antioxidant balance of the lung through oxidative stress is also thought to be an important contributor to the development of emphysema, an event that often occurs in tobacco smokers [11]. Specifically, cigarette smoke oxidatively damages the alpha-1 antitrypsin inhibitor, a compound that inhibits the activity of the enzyme elastase; elastase activity substantially increases, causing a breakdown in alveolar elastin and leading to impaired lung function [30]. While VAD contributes to conditions in the rat lung which increase the risk for emphysema, cigarette smoke inhalation has been shown to deplete lung vitamin A content, further contributing to the development of emphysema [30]. Emphysema, along with chronic bronchitis, comprises a condition known as COPD (Chronic Obstructive Pulmonary Disease); importantly, COPD is thought to be one of the most significant risk factors in lung cancer development and smokers who develop COPD have a greater risk of lung carcinogenesis compared to smokers who do not develop COPD [31, 11]. Thus, a relationship between cigarette smoke inhalation and VAD may account for a significant portion of emphysema development in animal models, while also impacting the risk of COPD development and, subsequently, lung cancer.

3. Lung Cancer and Vitamin A

3.1. Cancer and Retinoids: A Brief Overview

Many fruits and vegetables like sweet potatoes, butternut squash, spinach, oranges, and carrots are high in provitamin A carotenoids [4]. Epidemiological evidence shows increased fruit/vegetable intake to be associated with decreased risk of some cancers; High fruit and vegetable intakes are associated with a lower risk of colon cancer [32]. Increased consumption of fruits and vegetables is associated with a reduced risk of extrahepatic bile duct cancer and a small reduction in prostate cancer risk has been shown with higher fruit intake [33, 34]. Additionally, increased consumption of Vitamin C-rich vegetables was shown to decrease the risk of prostate cancer [35]. Amongst the elderly, supplemented vitamins A and C were shown to reduce colon cancer risk in women [36]. In men, decreased risk of renal cell cancer was associated with consumption of fruits and vegetables while intakes of vitamin A, C, and carotenoids from food were inversely associated with renal cell cancer risk [37].

As retinoic acid (RA), vitamin A has shown therapeutic promise in treating a form of cancer known as acute promyelocytic leukemia (APL) [38]. In combination with chemotherapeutic compounds such as arsenic trioxide (ATO), ATRA (all-trans retinoic acid) has been shown to lead to complete remission in 93% of patients and 5-year survival rates close to 100% [39]. Such effects have been observed to occur independently of ATO as well; Huang et al (1988) showed that ATRA (All-trans-retinoic acid) administration induced remission in 24 out of 24 cases of APL [40]. Evidence for ATRA having a protective role in other forms of cancer, however, is largely limited to in vitro

studies; ATRA has been shown to inhibit the growth of colorectal cancer cells and arrest the growth of cancer cells in prostate cancer cell culture [41].

Clinical trials investigating RA's effects on other forms of cancer have been less fruitful, due to a limited number of studies on the topic [42]. In vivo, RA's effectiveness on several different cancers, such as lung cancer, are mixed; this is likely due to RA's accelerated degradation with continuous administration [43]. Recent research has shown, however, that lipid-encapsulated RA formulas exhibit promising therapeutic benefits in treating lung cancer, compared to RA taken alone. In rats injected with lung cancer cell lines, ATRA encapsulated in a liposome consisting of phosphatidylcholine and cholesterol increased lifespan and reduced tumor module formation, compared to rats injected with ATRA only. Further, lung tissue of the rats given ATRA suspended in the liposome had significantly greater concentrations of ATRA; the authors of this study posited that the corresponding increase in ATRA in lung tissue could have been a product of increased transcription of RARs, some of which have been shown to have anti-cancer properties, such as RAR-Beta [44].

3.2. Dietary Vitamin A Intake, Supplemental Retinoid Intake, and Lung Cancer Risk

Outside of RA, several meta-analyses and studies have documented protective effects of fruit and vegetable intake, as well as dietary vitamin A/pro-vitamin A carotenoids on lung cancer risk. In a study examining the effects of individual Dietary Inflammatory Index (DII) and the Mediterranean Diet Score (MDS) on lung cancer risk, it was found that a higher MDS was inversely associated with lung cancer risk. A Mediterranean diet is often rich in vitamin A and pro-vitamin A carotenoids [26, 45]. Similarly, it has been shown that several dietary antioxidants, most notably vitamin C and pro-vitamin A

carotenoids, exert a protective effect against lung cancer in both moderate and heavy smokers [46]. A meta-analysis of 18 studies investigating fruit and vegetable consumption and comparative lung cancer risk found similar results; higher intakes of fruits and vegetables were shown to contribute an 8-18% lower risk of lung cancer in smokers [47]. Similarly, in a meta-analysis pooling 19 studies, it was found that higher intakes of foods containing beta-carotene and vitamin A were associated with decreased lung cancer risk [48].

While dietary intakes of vitamin A and/or antioxidant-rich fruits and vegetables may reduce the risk of lung cancer in both smokers and non-smokers, another important source of vitamin A in the American diet is cow's milk. While milk consumption in the US has decreased over the past few decades, the average American in 2018 drank around 17 gallons a year [49]. Thus, it is likely that milk is a substantial contributor to overall vitamin A sufficiency in the American diet. Generally, intakes of milk and dairy products are associated with improved health/disease outcomes [50]. Interestingly, evidence for milk consumption and lung cancer risk point to protective effects in some cases, and increased risk in other cases. In Sweden, a case-control study in patients with suspected lung cancer found an overall protective effect of vegetable intake but an increased risk of lung cancer with increasing milk consumption [51]. Investigators in the Guangzhou Biobank Cohort Study found that in individuals 50 years or older, moderate milk consumption was associated with lower cardiovascular disease mortality but high milk consumption was associated with a higher risk of total cancer mortality [52]. A hospital-based case-control study analyzing dietary habits of lung cancer patients found a 2-fold increase in lung cancer risk in patients who reported consuming whole milk

three or more times daily, compared to those who never drink milk. In this same study, a reduction in lung cancer risk was noted in individuals who consumed 2% or reduced-fat milk [53]. Compared to the observed protective effect of lung cancer in individuals who consume large amounts of fruits and vegetables, the effect of milk consumption on lung cancer risk is far less clear.

Outside of diet-based retinoid consumption, investigators have studied the effects of supplemental retinoids on lung cancer risk since the 1990s, with results pointing to a pronounced increase in lung cancer risk/incidence in smokers [11, 54, 55]. In ferrets exposed to cigarette smoke, supplemental carotenoids have been shown to increase lung cancer risk [55]. In the Beta-Carotene and Retinol Efficacy Trial, 18,314 smokers, former smokers, and workers exposed to asbestos were given a combination of supplemental beta-carotene and retinyl palmitate. The risk of lung cancer in those receiving supplements was drastically larger than those receiving the placebo; as such, the trial was stopped 21 months early [54]. As this result has been reproduced in several other large, interventional trials, the American Institute for Cancer Research (AICR) has labeled beta-carotene supplement use in smokers and former smokers as a cause of lung cancer [11]. The effects of supplemental beta-carotene on lung cancer risk have been hypothesized to be the product of an interaction between the supplement and cigarette smoke; a pro-oxidant effect of the cigarette smoke may lead to a free radical-induced transformation of beta-carotene to a structurally similar, yet biologically inert retinoid [11, 55]. Of note,

such an effect does not explain the increased risk of lung cancer observed in former smokers receiving beta-carotene supplements.

3.3. VAD, Cigarette Smoke, and Lung Cancer

In the United States alone, lung cancer is the second most common form of cancer in both men and women, with a combined 228,150 new cases appearing in 2019. It's also the most lethal form of cancer in both males and females, with estimated mortality rates of 24% and 23%, respectively; non-specific symptoms associated with cancer may contribute to this high mortality rate [56, 57]. The pathogenesis of lung cancer is complex, with roots in both genetics and environmental toxins. Furthering this complexity, the type of lung cancer may differ based on exposure to a specific carcinogen [58]. Of known risk factors for lung cancer development, tobacco smoke inhalation is the most prominent with an estimated 90% of new cases of lung cancer in some countries attributable to it [59].

The relationship between cigarette smoke and VAD is clearer than that of VAD and lung cancer risk. Almost three decades ago, researchers showed that benzopyrene-enriched rat feed induced a localized depletion of retinol in both the lung and the liver but not the serum of rats, pointing towards a potential role of vitamin A in protecting against lung cancer [60]. Later, researchers showed marked damage to both the liver and lungs when VAD was induced through the diet. In this model, the damage to the lungs mirrored that of emphysema [29]. As discussed in the previous section, emphysema in combination with chronic bronchitis results in a condition known as COPD. COPD is a significant risk factor for the development of lung cancer [31]. Subsequently, researchers showed that lung

retinoic acid content is depleted in rats exposed to cigarette smoke, with the tracheal tissue developing precancerous lesions [30].

Though cigarette smoke contributes to lung VAD it also contributes to cancer risk through the expression of proteins related to both vitamin A metabolism and cancer; lung retinoic acid content was depleted when rats were exposed to cigarette smoke which was correlated with a decrease in the expression of Retinoic Acid Receptor Beta (RAR-Beta), a protein implicated in decreased cancer risk [61-63]. Furthermore, many proteins associated with cancer growth were upregulated. Retinoic acid given to rats has been shown to down-regulate many proteins associated with cancer, even in the presence of cigarette smoke [63].

Independent of cigarette smoke, emerging evidence points to VAD increasing the expression of proteins linked to EMT (Epithelial-Mesenchymal Transition), a key stage in the progression and severity of respiratory pathologies such as COPD, pulmonary fibrosis, and lung cancer [64, 11]. Broadly, EMT occurs when epithelial cells lose proteins specific to their phenotype and gain proteins similar to a mesenchymal cell phenotype. While this transformation is characteristic of normal physiological processes such as those that occur during embryonic development and inflammation, its occurrence during lung carcinogenesis reduces the adherence capacity of cancer cells and increases the likelihood of metastasis to underlying tissues [64, 65]. The Notch signaling pathway is intimately involved in EMT progression, particularly as it pertains to the cell-specific transition to a mesenchymal phenotype in the tumor microenvironment [66]. Key to the involvement of Notch in this process is its interaction with Transforming Growth Factor-B1. Interestingly, VAD has been shown to increase the expression of Transforming

Growth Factor-B1 in animal models, pointing to a potential downstream relationship between VAD and EMT [67]. Down-regulation of proteins involved with cell-cell junction integrity and subsequent dissociation of these cell-cell junctions is a hallmark of the beginning stages of EMT. One such protein is E-cadherin, a protein that is also downregulated in cellular models of VAD [11, 64]. ATRA has been shown to activate E-cadherin expression as well as downregulate levels of Vimentin and Fibronectin in human colon carcinoma cells, pointing to a potential therapeutic mechanism for inhibiting EMT in certain forms of cancer [68]. Overall, the relationship between VAD and the EMT in lung cancer cells represents a promising avenue of research for the interaction between VAD and lung carcinogenesis.

4. References for Review of Literature

1. Das BC, Thapa P, Karki R, Das S, Mahapatra S, Liu TC, Torregroza I, Wallace DP, Kambhampati S, Van Veldhuizen P, Verma A, Ray SK, Evans T. Retinoic acid signaling pathways in development and diseases. Bioorg Med Chem. 2014 Jan 15;22(2):673-83. doi: 10.1016/j.bmc.2013.11.025. Epub 2013 Nov 22. PMID: 24393720; PMCID: PMC4447240.
2. McLaren DS, Kraemer K. Vitamin A in nature. World Rev Nutr Diet. 2012;103:7-17. doi: 10.1159/000258434. Epub 2012 Aug 27. PMID: 23008032.
3. Clagett-Dame M, Knutson D. Vitamin A in reproduction and development. Nutrients. 2011 Apr;3(4):385-428. doi: 10.3390/nu3040385. Epub 2011 Mar 29. PMID: 22254103; PMCID: PMC3257687.
4. WHO. Global prevalence of vitamin A deficiency in populations at risk 1995–2005. WHO Global Database on Vitamin A Deficiency. Geneva, World Health Organization, 2009.
5. Noy N, Slosberg E, Scarlata S. Interactions of retinol with binding proteins: studies with retinol-binding protein and with transthyretin. Biochemistry. 1992 Nov 17;31(45):11118-24. doi: 10.1021/bi00160a023. PMID: 1445851.
6. MacDonald PN, Ong DE. A lecithin:retinol acyltransferase activity in human and rat liver. Biochem Biophys Res Commun. 1988 Oct 14;156(1):157-63. doi: 10.1016/s0006-291x(88)80818-0. PMID: 3178828.
7. Blaner WS, O'Byrne SM, Wongsiriroj N, Kluwe J, D'Ambrosio DM, Jiang H, Schwabe RF, Hillman EM, Piantedosi R, Libien J. Hepatic stellate cell lipid droplets: a specialized lipid droplet for retinoid storage. Biochim Biophys Acta. 2009 Jun;1791(6):467-73. doi:

10.1016/j.bbalip.2008.11.001. Epub 2008 Nov 24. PMID: 19071229; PMCID: PMC2719539.

8. Blomhoff R, Blomhoff HK. Overview of retinoid metabolism and function. J Neurobiol. 2006 Jun;66(7):606-30. doi: 10.1002/neu.20242. PMID: 16688755.
9. Wei LN. Retinoid receptors and their coregulators. Annu Rev Pharmacol Toxicol. 2003;43:47-72. doi: 10.1146/annurev.pharmtox.43.100901.140301. Epub 2002 Jan 10. PMID: 12142470.
10. Bastien J, Rochette-Egly C. Nuclear retinoid receptors and the transcription of retinoid-target genes. Gene. 2004 Mar 17;328:1-16. doi: 10.1016/j.gene.2003.12.005. PMID: 15019979.
11. Timoneda J, Rodríguez-Fernández L, Zaragozá R, Marín MP, Cabezuelo MT, Torres L, Viña JR, Barber T. Vitamin A Deficiency and the Lung. Nutrients. 2018 Aug 21;10(9):1132. doi: 10.3390/nu10091132. PMID: 30134568; PMCID: PMC6164133.
12. Sommer A, Davidson FR; Annecy Accords. Assessment and control of vitamin A deficiency: the Annecy Accords. J Nutr. 2002 Sep;132(9 Suppl):2845S-2850S. doi: 10.1093/jn/132.9.2845S. PMID: 12221259.
13. WILSON JG, ROTH CB, WARKANY J. An analysis of the syndrome of malformations induced by maternal vitamin A deficiency. Effects of restoration of vitamin A at various times during gestation. Am J Anat. 1953 Mar;92(2):189-217. doi: 10.1002/aja.1000920202. PMID: 13030424.
14. El-Khashab EK, Hamdy AM, Maher KM, Fouad MA, Abbas GZ. Effect of maternal vitamin A deficiency during pregnancy on neonatal kidney size. J Perinat Med. 2013 Mar;41(2):199-203. doi: 10.1515/jpm-2012-0026. PMID: 23093302.

15. di Masi A, Leboffe L, De Marinis E, Pagano F, Cicconi L, Rochette-Egly C, Lo-Coco F, Ascenzi P, Nervi C. Retinoic acid receptors: from molecular mechanisms to cancer therapy. Mol Aspects Med. 2015 Feb;41:1-115. doi: 10.1016/j.mam.2014.12.003. Epub 2014 Dec 25. PMID: 25543955.
16. Maden M, Hind M. Retinoic acid in alveolar development, maintenance and regeneration. Philos Trans R Soc Lond B Biol Sci. 2004 May 29;359(1445):799-808. doi: 10.1098/rstb.2004.1470. PMID: 15293808; PMCID: PMC1693372.
17. Yun EJ, Lorizio W, Seedorf G, Abman SH, Vu TH. VEGF and endothelium-derived retinoic acid regulate lung vascular and alveolar development. Am J Physiol Lung Cell Mol Physiol. 2016 Feb 15;310(4):L287-98. doi: 10.1152/ajplung.00229.2015. Epub 2015 Nov 13. PMID: 26566904; PMCID: PMC4754906.
18. Massaro GD, Massaro D. Retinoic acid treatment partially rescues failed septation in rats and in mice. Am J Physiol Lung Cell Mol Physiol. 2000 May;278(5):L955-60. doi: 10.1152/ajplung.2000.278.5.L955. PMID: 10781425.
19. Wei H, Huang HM, Li TY, Qu P, Liu YX, Chen J. Marginal vitamin A deficiency affects lung maturation in rats from prenatal to adult stage. J Nutr Sci Vitaminol (Tokyo). 2009 Jun;55(3):208-14. doi: 10.3177/jnsv.55.208. PMID: 19602828.
20. Antipatis C, Grant G, Ashworth CJ. Moderate maternal vitamin A deficiency affects perinatal organ growth and development in rats. Br J Nutr. 2000 Jul;84(1):125-32. PMID: 10961169.
21. Christian P, Stewart CP. Maternal micronutrient deficiency, fetal development, and the risk of chronic disease. J Nutr. 2010 Mar;140(3):437-45. doi: 10.3945/jn.109.116327. Epub 2010 Jan 13. PMID: 20071652.

22. Shi W, Bellusci S, Warburton D. Lung development and adult lung diseases. Chest. 2007 Aug;132(2):651-6. doi: 10.1378/chest.06-2663. PMID: 17699136.

23. Brown CC, Noelle RJ. Seeing through the dark: New insights into the immune regulatory functions of vitamin A. Eur J Immunol. 2015 May;45(5):1287-95. doi: 10.1002/eji.201344398. PMID: 25808452; PMCID: PMC4426035.

24. Sommer A, Katz J, Tarwotjo I. Increased risk of respiratory disease and diarrhea in children with preexisting mild vitamin A deficiency. Am J Clin Nutr. 1984 Nov;40(5):1090-5. doi: 10.1093/ajcn/40.5.1090. PMID: 6496388.

25. Karim T, Muhit M, Khandaker G. Interventions to prevent respiratory diseases - Nutrition and the developing world. Paediatr Respir Rev. 2017 Mar;22:31-37. doi: 10.1016/j.prrv.2016.09.003. Epub 2016 Sep 28. PMID: 27793738.

26. Garcia-Larsen V, Del Giacco SR, Moreira A, Bonini M, Charles D, Reeves T, Carlsen KH, Haahtela T, Bonini S, Fonseca J, Agache I, Papadopoulos NG, Delgado L. Asthma and dietary intake: an overview of systematic reviews. Allergy. 2016 Apr;71(4):433-42. doi: 10.1111/all.12800. Epub 2016 Jan 19. PMID: 26505989.

27. Luo ZX, Liu EM, Luo J, Li FR, Li SB, Zeng FQ, Qu P, Fu Z, Li TY. Vitamin A deficiency and wheezing. World J Pediatr. 2010 Feb;6(1):81-4. doi: 10.1007/s12519-010-0012-7. Epub 2010 Feb 9. PMID: 20143217.

28. McGowan SE, Holmes AJ, Smith J. Retinoic acid reverses the airway hyperresponsiveness but not the parenchymal defect that is associated with vitamin A deficiency. Am J Physiol Lung Cell Mol Physiol. 2004 Feb;286(2):L437-44. doi: 10.1152/ajplung.00158.2003. PMID: 14711804.

29. Baybutt RC, Hu L, Molteni A. Vitamin A deficiency injures lung and liver parenchyma and impairs function of rat type II pneumocytes. J Nutr. 2000 May;130(5):1159-65. doi: 10.1093/jn/130.5.1159. PMID: 10801913.

30. Li T, Molteni A, Latkovich P, Castellani W, Baybutt RC. Vitamin A depletion induced by cigarette smoke is associated with the development of emphysema in rats. J Nutr. 2003 Aug;133(8):2629-34. doi: 10.1093/jn/133.8.2629. PMID: 12888649.

31. Sekine Y, Katsura H, Koh E, Hiroshima K, Fujisawa T. Early detection of COPD is important for lung cancer surveillance. Eur Respir J. 2012 May;39(5):1230-40. doi: 10.1183/09031936.00126011. Epub 2011 Nov 16. PMID: 22088970.

32. Leenders M, Siersema PD, Overvad K, Tjønneland A, Olsen A, Boutron-Ruault MC, Bastide N, Fagherazzi G, Katzke V, Kühn T, Boeing H, Aleksandrova K, Trichopoulou A, Lagiou P, Klinaki E, Masala G, Grioni S, Santucci De Magistris M, Tumino R, Ricceri F, Peeters PH, Lund E, Skeie G, Weiderpass E, Quirós JR, Agudo A, Sánchez MJ, Dorronsoro M, Navarro C, Ardanaz E, Ohlsson B, Jirström K, Van Guelpen B, Wennberg M, Khaw KT, Wareham N, Key TJ, Romieu I, Huybrechts I, Cross AJ, Murphy N, Riboli E, Bueno-de-Mesquita HB. Subtypes of fruit and vegetables, variety in consumption and risk of colon and rectal cancer in the European Prospective Investigation into Cancer and Nutrition. Int J Cancer. 2015 Dec 1;137(11):2705-14. doi: 10.1002/ijc.29640. Epub 2015 Jul 16. PMID: 26077137.

33. Makiuchi T, Sobue T, Kitamura T, Ishihara J, Sawada N, Iwasaki M, Sasazuki S, Yamaji T, Shimazu T, Tsugane S. The relationship between vegetable/fruit consumption and gallbladder/bile duct cancer: A population-based cohort study in Japan. Int J Cancer.

2017 Mar 1;140(5):1009-1019. doi: 10.1002/ijc.30492. Epub 2016 Nov 29. PMID: 27798952.

34. Perez-Cornago A, Travis RC, Appleby PN, Tsilidis KK, Tjønneland A, Olsen A, Overvad K, Katzke V, Kühn T, Trichopoulou A, Peppa E, Kritikou M, Sieri S, Palli D, Sacerdote C, Tumino R, Bueno-de-Mesquita HBA, Agudo A, Larrañaga N, Molina-Portillo E, Ardanaz E, Chirlaque MD, Lasheras C, Stattin P, Wennberg M, Drake I, Malm J, Schmidt JA, Khaw KT, Gunter M, Freisling H, Huybrechts I, Aune D, Cross AJ, Riboli E, Key TJ. Fruit and vegetable intake and prostate cancer risk in the European Prospective Investigation into Cancer and Nutrition (EPIC). Int J Cancer. 2017 Jul 15;141(2):287-297. doi: 10.1002/ijc.30741. Epub 2017 May 15. PMID: 28419475; PMCID: PMC5488166.
35. Ambrosini GL, de Klerk NH, Fritschi L, Mackerras D, Musk B. Fruit, vegetable, vitamin A intakes, and prostate cancer risk. Prostate Cancer Prostatic Dis. 2008;11(1):61-6. doi: 10.1038/sj.pcan.4500979. Epub 2007 May 22. PMID: 17519926.
36. Shibata A, Paganini-Hill A, Ross RK, Henderson BE. Intake of vegetables, fruits, beta-carotene, vitamin C and vitamin supplements and cancer incidence among the elderly: a prospective study. Br J Cancer. 1992 Oct;66(4):673-9. doi: 10.1038/bjc.1992.336. PMID: 1419605; PMCID: PMC1977409.
37. Lee JE, Giovannucci E, Smith-Warner SA, Spiegelman D, Willett WC, Curhan GC. Intakes of fruits, vegetables, vitamins A, C, and E, and carotenoids and risk of renal cell cancer. Cancer Epidemiol Biomarkers Prev. 2006 Dec;15(12):2445-52. doi: 10.1158/1055-9965.EPI-06-0553. PMID: 17164369.

38. Schenk T, Stengel S, Zelent A. Unlocking the potential of retinoic acid in anticancer therapy. Br J Cancer. 2014 Nov 25;111(11):2039-45. doi: 10.1038/bjc.2014.412. Epub 2014 Nov 20. PMID: 25412233; PMCID: PMC4260020.

39. Lo-Coco F, Avvisati G, Vignetti M, Thiede C, Orlando SM, Iacobelli S, Ferrara F, Fazi P, Cicconi L, Di Bona E, Specchia G, Sica S, Divona M, Levis A, Fiedler W, Cerqui E, Breccia M, Fioritoni G, Salih HR, Cazzola M, Melillo L, Carella AM, Brandts CH, Morra E, von Lilienfeld-Toal M, Hertenstein B, Wattad M, Lübbert M, Hänel M, Schmitz N, Link H, Kropp MG, Rambaldi A, La Nasa G, Luppi M, Ciceri F, Finizio O, Venditti A, Fabbiano F, Döhner K, Sauer M, Ganser A, Amadori S, Mandelli F, Döhner H, Ehninger G, Schlenk RF, Platzbecker U; Gruppo Italiano Malattie Ematologiche dell'Adulto; German-Austrian Acute Myeloid Leukemia Study Group; Study Alliance Leukemia. Retinoic acid and arsenic trioxide for acute promyelocytic leukemia. N Engl J Med. 2013 Jul 11;369(2):111-21. doi: 10.1056/NEJMoa1300874. PMID: 23841729.

40. Lo-Coco F, Avvisati G, Vignetti M, Thiede C, Orlando SM, Iacobelli S, Ferrara F, Fazi P, Cicconi L, Di Bona E, Specchia G, Sica S, Divona M, Levis A, Fiedler W, Cerqui E, Breccia M, Fioritoni G, Salih HR, Cazzola M, Melillo L, Carella AM, Brandts CH, Morra E, von Lilienfeld-Toal M, Hertenstein B, Wattad M, Lübbert M, Hänel M, Schmitz N, Link H, Kropp MG, Rambaldi A, La Nasa G, Luppi M, Ciceri F, Finizio O, Venditti A, Fabbiano F, Döhner K, Sauer M, Ganser A, Amadori S, Mandelli F, Döhner H, Ehninger G, Schlenk RF, Platzbecker U; Gruppo Italiano Malattie Ematologiche dell'Adulto; German-Austrian Acute Myeloid Leukemia Study Group; Study Alliance Leukemia. Retinoic acid and arsenic trioxide for acute promyelocytic leukemia. N Engl J Med. 2013 Jul 11;369(2):111-21. doi: 10.1056/NEJMoa1300874. PMID: 23841729.

41. Liu W, Song Y, Zhang C, Gao P, Huang B, Yang J. The protective role of all-transretinoic acid (ATRA) against colorectal cancer development is achieved via increasing miR-3666 expression and decreasing E2F7 expression. Biomed Pharmacother. 2018 Aug;104:94-101. doi: 10.1016/j.biopha.2018.05.015. Epub 2018 May 14. PMID: 29772445.
42. Garattini E, Bolis M, Garattini SK, Fratelli M, Centritto F, Paroni G, Gianni' M, Zanetti A, Pagani A, Fisher JN, Zambelli A, Terao M. Retinoids and breast cancer: from basic studies to the clinic and back again. Cancer Treat Rev. 2014 Jul;40(6):739-49. doi: 10.1016/j.ctrv.2014.01.001. Epub 2014 Jan 18. PMID: 24480385.
43. Viswanathan S, Berlin Grace VM, Danisha JP. Enhancement of tumor suppressor RAR-β protein expression by cationic liposomal-ATRA treatment in benzo(a)pyrene-induced lung cancer mice model. Naunyn Schmiedebergs Arch Pharmacol. 2019 Apr;392(4):415-426. doi: 10.1007/s00210-018-01598-8. Epub 2018 Dec 12. PMID: 30539216.
44. Siddikuzzaman, Grace VM. Inhibition of metastatic lung cancer in C57BL/6 mice by liposome encapsulated all trans retinoic acid (ATRA). Int Immunopharmacol. 2012 Dec;14(4):570-9. doi: 10.1016/j.intimp.2012.09.008. Epub 2012 Sep 27. PMID: 23021983.
45. Shareck M, Rousseau MC, Koushik A, Siemiatycki J, Parent ME. Inverse Association between Dietary Intake of Selected Carotenoids and Vitamin C and Risk of Lung Cancer. Front Oncol. 2017 Feb 28;7:23. doi: 10.3389/fonc.2017.00023. PMID: 28293540; PMCID: PMC5328985.
46. Hodge AM, Bassett JK, Shivappa N, Hébert JR, English DR, Giles GG, Severi G. Dietary inflammatory index, Mediterranean diet score, and lung cancer: a prospective

study. Cancer Causes Control. 2016 Jul;27(7):907-17. doi: 10.1007/s10552-016-0770-1. Epub 2016 Jun 13. PMID: 27294725; PMCID: PMC5550291.

47. Vieira AR, Abar L, Vingeliene S, Chan DS, Aune D, Navarro-Rosenblatt D, Stevens C, Greenwood D, Norat T. Fruits, vegetables and lung cancer risk: a systematic review and meta-analysis. Ann Oncol. 2016 Jan;27(1):81-96. doi: 10.1093/annonc/mdv381. Epub 2015 Sep 14. PMID: 26371287.
48. Yu N, Su X, Wang Z, Dai B, Kang J. Association of Dietary Vitamin A and β-Carotene Intake with the Risk of Lung Cancer: A Meta-Analysis of 19 Publications. Nutrients. 2015 Nov 11;7(11):9309-24. doi: 10.3390/nu7115463. PMID: 26569298; PMCID: PMC4663591.
49. Cessna J, Teran A. Diary Products: Per Capita Consumption, United States (Annual) [Internet]. United States Department of Agriculture; 2020 September [Accessed 17 June 2021] https://www.ers.usda.gov/data-products/dairy-data/.
50. Thorning TK, Raben A, Tholstrup T, Soedamah-Muthu SS, Givens I, Astrup A. Milk and dairy products: good or bad for human health? An assessment of the totality of scientific evidence. Food Nutr Res. 2016 Nov 22;60:32527. doi: 10.3402/fnr.v60.32527. PMID: 27882862; PMCID: PMC5122229.
51. Axelsson G, Liljeqvist T, Andersson L, Bergman B, Rylander R. Dietary factors and lung cancer among men in west Sweden. Int J Epidemiol. 1996 Feb;25(1):32-9. doi: 10.1093/ije/25.1.32. PMID: 8666501.
52. Wang XJ, Jiang CQ, Zhang WS, Zhu F, Jin YL, Woo J, Cheng KK, Lam TH, Xu L. Milk consumption and risk of mortality from all-cause, cardiovascular disease and cancer in

older people. Clin Nutr. 2020 Nov;39(11):3442-3451. doi: 10.1016/j.clnu.2020.03.003. Epub 2020 Mar 12. PMID: 32229169.

53. Mettlin C. Milk drinking, other beverage habits, and lung cancer risk. Int J Cancer. 1989 Apr 15;43(4):608-12. doi: 10.1002/ijc.2910430412. PMID: 2703270.
54. Wang XD, Liu C, Bronson RT, Smith DE, Krinsky NI, Russell M. Retinoid signaling and activator protein-1 expression in ferrets given beta-carotene supplements and exposed to tobacco smoke. J Natl Cancer Inst. 1999 Jan 6;91(1):60-6. doi: 10.1093/jnci/91.1.60. PMID: 9890171.
55. Omenn GS, Goodman GE, Thornquist MD, Balmes J, Cullen MR, Glass A, Keogh JP, Meyskens FL, Valanis B, Williams JH, Barnhart S, Hammar S. Effects of a combination of beta carotene and vitamin A on lung cancer and cardiovascular disease. N Engl J Med. 1996 May 2;334(18):1150-5. doi: 10.1056/NEJM199605023341802. PMID: 8602180
56. Xing PY, Zhu YX, Wang L, Hui ZG, Liu SM, Ren JS, Zhang Y, Song Y, Liu CC, Huang YC, Liao XZ, Xing XJ, Wang DB, Yang L, Du LB, Liu YQ, Zhang YZ, Liu YY, Wei DH, Zhang K, Shi JF, Qiao YL, Chen WQ, Li JL, Dai M; LuCCRES Group. What are the clinical symptoms and physical signs for non-small cell lung cancer before diagnosis is made? A nation-wide multicenter 10-year retrospective study in China. Cancer Med. 2019 Jul;8(8):4055-4069. doi: 10.1002/cam4.2256. Epub 2019 May 31. PMID: 31150167; PMCID: PMC6639195.
57. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. CA Cancer J Clin. 2019 Jan;69(1):7-34. doi: 10.3322/caac.21551. Epub 2019 Jan 8. PMID: 30620402.
58. Richiardi L, Boffetta P, Simonato L, Forastiere F, Zambon P, Fortes C, Gaborieau V, Merletti F. Occupational risk factors for lung cancer in men and women: a population-

based case-control study in Italy. Cancer Causes Control. 2004 Apr;15(3):285-94. doi: 10.1023/B:CACO.0000024223.91059.ed. PMID: 15090723.

59. Pesch B, Kendzia B, Gustavsson P, Jöckel KH, Johnen G, Pohlabeln H, Olsson A, Ahrens W, Gross IM, Brüske I, Wichmann HE, Merletti F, Richiardi L, Simonato L, Fortes C, Siemiatycki J, Parent ME, Consonni D, Landi MT, Caporaso N, Zaridze D, Cassidy A, Szeszenia-Dabrowska N, Rudnai P, Lissowska J, Stücker I, Fabianova E, Dumitru RS, Bencko V, Foretova L, Janout V, Rudin CM, Brennan P, Boffetta P, Straif K, Brüning T. Cigarette smoking and lung cancer--relative risk estimates for the major histological types from a pooled analysis of case-control studies. Int J Cancer. 2012 Sep 1;131(5):1210-9. doi: 10.1002/ijc.27339. Epub 2011 Dec 14. PMID: 22052329; PMCID: PMC3296911.
60. Edes TE, Gysbers DS. Carcinogen-induced tissue vitamin A depletion. Potential protective advantages of beta-carotene. Ann N Y Acad Sci. 1993 May 28;686:203-11; discussion 211-2. doi: 10.1111/j.1749-6632.1993.tb39176.x. PMID: 8512248.
61. Muñiz-Hernández S, Huerta-Yepez S, Hernández-Pedro N, Ramírez-Tirado LA, Aviles-Salas A, Maldonado A, Hernández-Cueto D, Baay-Guzmán G, Arrieta O. Association between nuclear expression of retinoic acid receptor alpha and beta and clinicopathological features and prognosis of advanced non-small cell lung cancer. Int J Clin Oncol. 2016 Dec;21(6):1051-1061. doi: 10.1007/s10147-016-1002-0. Epub 2016 Jun 15. PMID: 27306217.
62. Feng H, Zhang Z, Qing X, Wang X, Liang C, Liu D. Promoter methylation of APC and RAR-β genes as prognostic markers in non-small cell lung cancer (NSCLC). Exp Mol

Pathol. 2016 Feb;100(1):109-13. doi: 10.1016/j.yexmp.2015.12.005. Epub 2015 Dec 8. PMID: 26681652.

63. Xue Y, Harris E, Wang W, Baybutt RC. Vitamin A depletion induced by cigarette smoke is associated with an increase in lung cancer-related markers in rats. J Biomed Sci. 2015 Oct 14;22:84. doi: 10.1186/s12929-015-0189-0. PMID: 26462767; PMCID: PMC4605095.
64. Sung WJ, Kim H, Park KK. The biological role of epithelial-mesenchymal transition in lung cancer (Review). Oncol Rep. 2016 Sep;36(3):1199-206. doi: 10.3892/or.2016.4964. Epub 2016 Jul 21. PMID: 27460444
65. De Craene B, Berx G. Regulatory networks defining EMT during cancer initiation and progression. Nat Rev Cancer. 2013 Feb;13(2):97-110. doi: 10.1038/nrc3447. PMID: 23344542.
66. Koren A, Motaln H, Cufer T. Lung cancer stem cells: a biological and clinical perspective. Cell Oncol (Dordr). 2013 Jul;36(4):265-75. doi: 10.1007/s13402-013-0141-9. Epub 2013 Jul 31. PMID: 23900738.
67. Esteban-Pretel G, Marín MP, Renau-Piqueras J, Barber T, Timoneda J. Vitamin A deficiency alters rat lung alveolar basement membrane: reversibility by retinoic acid. J Nutr Biochem. 2010 Mar;21(3):227-36. doi: 10.1016/j.jnutbio.2008.12.007. Epub 2009 Mar 9. PMID: 19269151.
68. Woo YJ, Jang KL. All-trans retinoic acid activates E-cadherin expression via promoter hypomethylation in the human colon carcinoma HCT116 cells. Biochem Biophys Res Commun. 2012 Sep 7;425(4):944-9. doi: 10.1016/j.bbrc.2012.08.038. Epub 2012 Aug 15. PMID: 22910408.

CHAPTER 2: BODY OF THESIS

1. Introduction

Vitamin A plays a crucial role in cell/tissue growth and differentiation, cell signaling, reproduction, epithelial integrity, vision, and maintenance of the immune system, across all stages of life [1, 2]. Vitamin A Deficiency (VAD), particularly during periods when the body undergoes rapid growth, is detrimental because of its complex role in important bodily processes such as vision, epithelial integrity, and immune function; as such, VAD, especially in children, can lead to preventable blindness and increased risk of death from respiratory infections [3,4]. Impairments to the respiratory tract and associated organs are established consequences of VAD [5]. VAD has been shown to cause profound morphological changes to both the rat lung and liver, impairing pneumocyte function, surfactant production, and promoting inflammation [6]. Additionally, increased rates of respiratory infections, increased risk of asthmatic symptoms, and increased risk of emphysema are often observed in conditions where VAD is present [2, 7, 8]. Additionally, VAD may increase the risk of developing emphysema, COPD, and lung fibrosis, particularly in smokers [5].

Unlike pathologies such as emphysema, asthma, and lung fibrosis the involvement of VAD in the etiology of lung cancer is not as clear. Lung cancer is the second most common form of cancer in the US and is also the most lethal [9]. Inhalation of tobacco smoke is the most well-known risk factor for lung cancer development [10]. The development of lung cancer (lung carcinogenesis) in tobacco smokers is thought to be largely attributable to a combination of oxidative/free radical damage and DNA damage, particularly in the form of DNA adducts and their metabolites [11]. Over the past decades,

emerging evidence has pointed to an interaction between cigarette smoke and vitamin A in the lung; inhalation of cigarette smoke in animals has shown to induce VAD in the lungs and increase lung cancer risk [12]. Further, VAD induced by cigarette smoke has also been shown to lead to the formation of pre-cancerous, tracheal lesions and has been shown to deplete lung retinoic acid content while increasing cancer risk through the expression/suppression of certain cancer-related proteins [12, 13]. Independent of cigarette smoke, VAD alone is being investigated for its potential role in a process known as Epithelial-Mesenchymal Transition, or EMT [5]. EMT increases the likelihood of established cancer to invade underlying tissues; VAD impacts several proteins implicated in both signaling and structural components of the process such as the Notch signaling pathway, Transforming Growth Factor-B1, and e-Cadherin [14-17]. Combined with the established pathophysiology of cigarette smoke, VAD in pulmonary tissue has shown and is continuing to show evidence that may link it to a contributory role in lung carcinogenesis.

In contrast to VAD, the relationship between adequate dietary intakes of vitamin A and lung cancer has been studied extensively, albeit observationally. Recent epidemiological evidence indicates that diets rich in fruits and vegetables, especially those containing provitamin A carotenoids and other antioxidants, reduce the risk of lung cancer in the general population as well as heavy and moderate smokers [18, 19]. Elsewhere, recent meta-analyses showed reductions in lung cancer risk amongst current smokers, but not former and never smokers, as well as reductions in lung cancer risk with increasing intakes of beta-carotene from the diet in the general population [20, 21]. While intakes of fruits and vegetables rich in the provitamin A carotenoids and other antioxidant nutrients may

reduce lung cancer risk in smokers, intakes of fruits and vegetables amongst smokers are lower compared to non-smokers [22].

While cigarette smoke-induced VAD increases the risk for lung cancer in animal models, it is unclear as to what extent VAD contributes to the increased cancer risk in humans [23]. Though adequate dietary vitamin A is broadly protective against lung cancer, there is limited evidence available to determine what extent intakes of the vitamin lower than the minimal physiological requirements to prevent deficiency affect lung cancer risk. Thus, we sought to investigate this relationship through a retrospective cohort study derived from NHANES data. Our working hypothesis is that localized depletion of lung vitamin A by cigarette smoke may contribute to lung carcinogenesis through physiological changes as well as changes to the microcellular environment; consequently, insufficient/very low dietary vitamin A intake from food may expedite this process in individuals who smoke cigarettes. Contrarily, intake of abundant vitamin A from the diet may delay lung carcinogenesis in smokers.

Through a case-control design, we looked at individuals with lung cancer matched to cancer-free controls with similar histories of cigarette smoking. To our knowledge, no existing case-control studies have examined the relationship between dietary retinoid intakes and lung cancer risk, using smoking duration as the primary matching criteria. Further, we are unaware of any studies that have looked at lung cancer incidence and its association with dietary vitamin A intakes by classification of less than or greater than gender-specific, minimal physiological requirements. By using such criteria, we hoped to discover whether rates of vitamin A insufficiency differed significantly between case and control, given similar smoking history. Through this, we aimed to determine whether

intakes of retinoids lower than established minimum requirements could increase the risk of lung cancer.

2. Methods

2.1. Data Collection

Participants were selected from eight cycles (2003/2004 – 2017/2018) of the National Health and Nutrition Examination Survey (NHANES). A function of the National Center for Health and Statistics (NCHS), the NHANES is a cross-sectional survey designed to analyze trends in the health and nutritional status of adults and children across the US [24]. Information is collected through physical examinations, interviews, questionnaires, and laboratory analyses. Information collected includes, but is not limited to dietary intakes, laboratory values, smoking habits, demographic information, anthropometric measurements, medical history, medication/supplement use, and dental health. NHANES is designed to collect information that is nationally representative of the United States civilian (non-institutionalized) population. As such, data is collected through a four-stage sampling procedure each cycle, across 15 different national locations. New surveys are collected every 2 years and consist of approximately 10,000 individuals a survey. Each cycle, certain subsets of the population who are of public health interest are oversampled to produce estimates representative of the non-institutionalized, US population [25]. The NHANES protocol was approved by the NCHS Ethics Review Board. Further, participants had to provide signed, written consent to be a part of NHANES data collection. As our study was a secondary analysis using de-identified personal data, it did not require Institutional Review Board (IRB) review. All

data collected as part of our study is publicly available at https://www.cdc.gov/nchs/nhanes/.

2.2. Research Design

A case-control approach using pairwise, 1:1 matching of the case to control was used to investigate whether insufficient dietary vitamin A increases the risk for lung cancer and whether sufficient dietary intakes are protective against lung cancers in current smokers. Broadly, a case-control study design uses a preselected group of cases with an outcome of interest, along with a control group constructed similarly to the case group but without the outcome of interest. Following this, researchers can compare exposures across groups to determine whether certain risks are more common in one group versus the other [26]. Our case-control approach was advantageous for our research question because the number of lung cancer cases was low in our initial sample [26]. This approach allowed us to evaluate our target population efficiently, albeit retrospectively. Further, case-control approaches help assess the impact of multiple risk factors on a target outcome at once as well as potential factors that could decrease the risk of the outcome as well. Thus, our design allowed us to explore the impact of various factors on the outcome and assess their independent and collective contributions to lung cancer risk [26]. This helped us answer our research questions and determine the sample-specific etiology of lung cancer risk. For more details concerning the construction of the case and control groups, as well as details concerning our matching process, please see **Figure 1** and section 2.4,

"Exclusion Criteria and Group Selection". Specifics concerning variables of interest and our outcome can be found in Table 1 as well as section 2.3, "Variable Selection".

2.3. Variable Selection

Table 1 outlines variables of interest in our investigation, as well as cycle by cycle changes in question formatting or coding in the NHANES data files. **Table 2** outlines changes made to the coding or output of variables of interest, as well as transformations made to the variables. Variables of interest in our study were chosen for their importance/relationship to the primary outcome and our investigation of influential independent variables on this outcome.

Figure 1

Schematic for Exclusion Criteria, Case Assignment, Control Assignment, and Matching

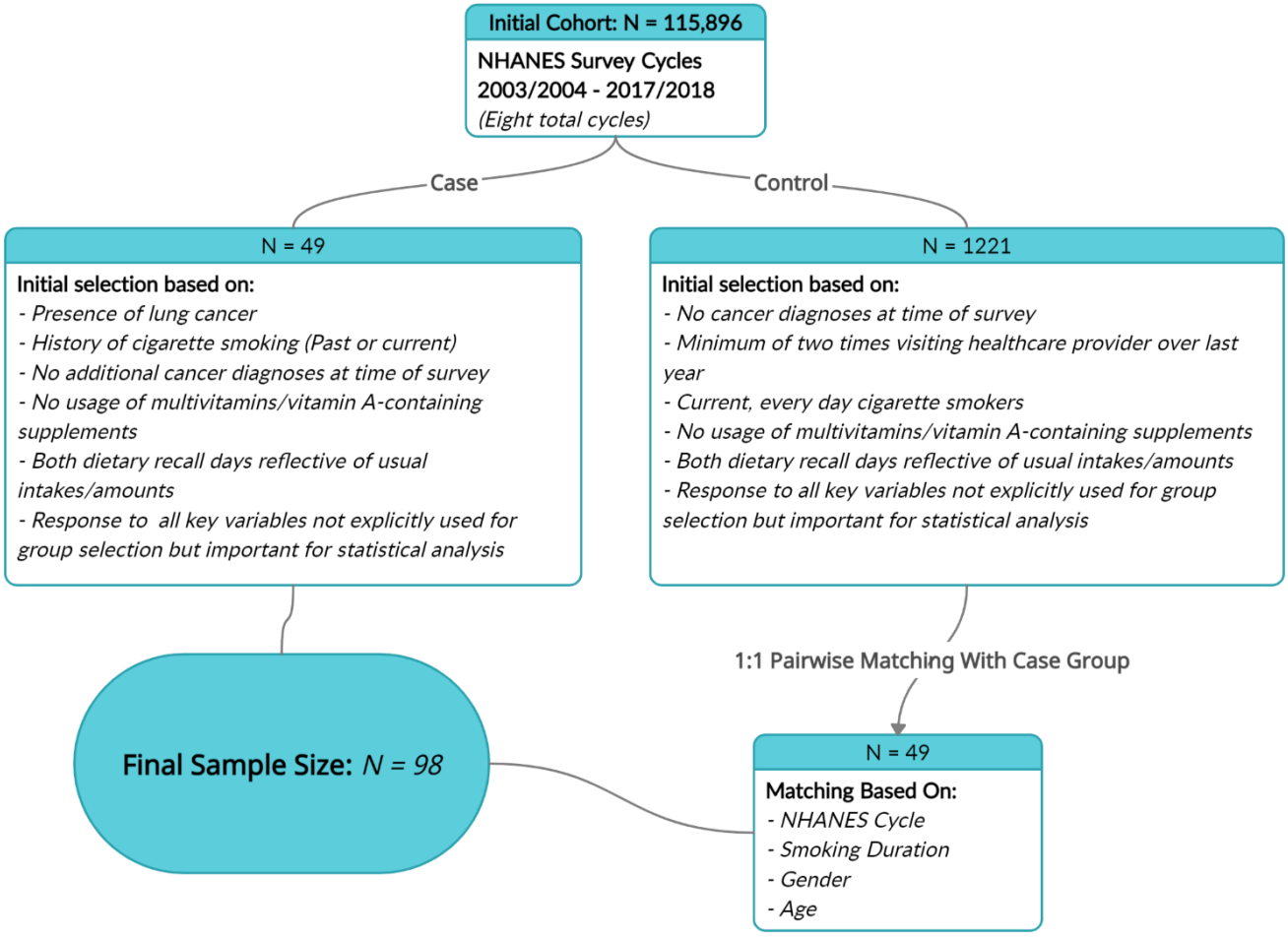

Table 1

Variables of Interest for Case/Control Establishment, Exclusion Criteria, and Statistical Analysis

Variable Name/Code	Variable Location Within NHANES Cycle File	Answer Choices	Cycle Changes
MCQ240N – "Age When Lung Cancer First Diagnosed"	Medical Conditions Data File	If answered, age of cancer diagnosis is provided	2017-2018 cycle: Instead of asking "Age When Lung Cancer First Diagnosed", CAPI asks interviewees about their 1st, 2nd, and 3rd cancer diagnoses, if applicable.
RIDAGEYR – "Age in Years at Screening"	Demographic Variables and Sample Weights	Numerical Answer Input	N/A
RIAGENDER – "Gender"	Demographic Variables and Sample Weights	Binary (Male/Female)	N/A
DMEDUC2 – "Education Level – Adults 20+"	Demographic Variables and Sample Weights	Ordinal (7 categories) based on level of education.	N/A

BMXWAIST – "Waist Circumference (cm)"	Body Measures	Numerical Answer Input	N/A
HUQ050 – "#Times Receive Healthcare Over Past Year"	Hospital Utilization & Access to Care	Ordinal (8 categories) based on the range of times reported by the participant.	N/A
DR1+2IFDCD – "USDA Food Code"	Dietary Interview – Individual Foods, First Day & Second Day	Food Codes were provided based on FNDDS data with the corresponding codebook; any recorded instances of milk consumption were noted.	Cycle to cycle changes in what certain codes encompass

DR1+2TVARA – "Vitamin A, RAE (mcg)" **DR1+2TACAR – "Alpha-carotene (mcg)"** **DR1+2TBCAR – "Beta-carotene (mcg)"** **DR1+2TCRYP – "Beta-cryptoxanthin (mcg)"** **DR1+2TRET – "Retinol (mcg)"** **DR1+2TP182 – "PFA 18:2 (Octadecadienoic) (gm)"** **DR1+2TP183 – "PFA 18:3 (Octadecatrienoic) (gm)"** **DR1+2TP205 – "PFA 20:5 (Eicosapentaenoic) (gm)"** **DR1+2TP226 – "PFA 22:6 (Docosahexaenoic) (gm)"**	Dietary Interview – Total Nutrient Intakes, First & Second Day	Individual nutrient intakes are provided numerically, based on 24-hr recall reports and corresponding food codes.	N/A
DR1+2300 – "Compare food consumed yesterday to usual"	Dietary Interview - Total Nutrient Intakes, First & Second Day	Whether or not intakes from 24-hr recalls are similar to the amount usually consumed	N/A

DSDSUPID - Supplement ID Number	Dietary Supplement Use 30 Day - Individual Dietary Supplements	Trained interviewers collect information on supplement usage from participants during the in-home interview portion of NHANES.	In 2003-2006 (two cycles), supplement ID codes and supplement names were located in separate files "Dietary Supplement Use 30-Day - File 2, Participant's Use of Supplements" and "Dietary Supplement Database - Product Information" and had to be matched before usage.
DBQ197 – "Past 30-day milk product consumption"	Diet Behavior & Nutrition	Participants choose a frequency of milk product consumption over the past 30 months, with choices from 0 (never) to 4 (varied)	N/A
SMD030 – "Age started smoking cigarettes regularly"	Smoking - Cigarette/Tobacco Use – Adult	Individuals list age and if never smoked regularly, "0"	After the 2003-2004 Cycle, the Title of the Dataset changed to "Smoking – Cigarette Use" to encompass all age groups of participants.

SMQ040 – "Do you now smoke cigarettes"	Smoking - Cigarette/Tobacco Use – Adult	Individuals choose between "Every day", "Some days", or "Not at All"	After the 2003-2004 Cycle, the Title of the Dataset changed to "Smoking – Cigarette Use" to encompass all age groups of participants.
SMQ050Q – "How long since quit smoking cigarettes" **&** **SMQ050U – "Unit of measure (day/week/month/year)"**	Smoking - Cigarette/Tobacco Use – Adult	Individuals choose a value and correspond it with a unit of measurement, "Days", "Weeks", "Months", "Years"	After the 2003-2004 Cycle, the Title of the Dataset changed to "Smoking – Cigarette Use" to encompass all age groups of participants.
SMD055 – "Age last smoked cigarettes regularly"	Smoking - Cigarette/Tobacco Use – Adult	Participants list a numerical value for this question.	After the 2003-2004 Cycle, the Title of the Dataset changed to "Smoking – Cigarette Use" to encompass all age groups of participants.
SMD057 – "# cigarettes smoked per day when quit"	Smoking - Cigarette/Tobacco Use - Adult	Participants list a numerical value for this question.	After the 2003-2004 Cycle, the Title of the Dataset changed to "Smoking – Cigarette Use" to encompass all age groups of participants.

SMD070 – "# cigarettes smoked per day now"	Smoking - Cigarette/Tobacco Use - Adult	Participants list a numerical value for this question.	After the 2003-2004 Cycle, the Title of the Dataset changed to "Smoking – Cigarette Use" to encompass all age groups of participants.
SMD410 – "Does anyone smoke in the home?"	Smoking - Household Smokers	Participants answer yes or no to this question.	N/A

Note: In their original format, dietary variables associated with the first recall day were indicated by "DR1_" followed by the specific food or nutrient. Dietary variables associated with the second recall day were indicated by "DR2_". For the sake of organization and space, we indicated any dietary variables which were recorded for both recall days as "DR1+2", followed by the original food/nutrient code. For example, the DR1+2TVARA – "Vitamin A, RAE (mcg)" variable (as shown in our table) was derived from DR1_TVARA and DR2_TVARA.

Table 2

Variables transformed or recoded for analysis

Original Variable(s)	Summary of Changes	New Variable
DR1+2TVARA – "Vitamin A, RAE (mcg)" **DR1+2TACAR – "Alpha-carotene (mcg)"** **DR1+2TBCAR – "Beta-carotene (mcg)"** **DR1+2TCRYP – "Beta-cryptoxanthin (mcg)"** **DR1+2TRET – "Retinol (mcg)"**	Summed two-day recall totals of each retinoid and provitamin A carotenoid together then divided the value by two to produce an average for each participant.	"Mean Two-Day Total Vitamin A Intakes" "Mean Two-Day Total Retinol Intakes" "Mean Two-Day Total Beta Carotene Intakes" "Mean Two-Day Total Alpha Carotene Intakes" "Mean Two-Day Total Beta Cryptoxanthin Intakes"
"Mean Two-Day Total Beta Carotene Intakes" **"Mean Two-Day Total Alpha Carotene Intakes"** **"Mean Two-Day Total Beta Cryptoxanthin Intakes"**	Produced a total carotenoid intake variable by adding all intakes of carotenoids for each participant together.	"Mean Total Two-Day Carotenoid Intakes"

DR1+2TP182 – "PFA 18:2 (Octadecadienoic) (gm)" **DR1+2TP183 – "PFA 18:3 (Octadecatrienoic) (gm)"** **DR1+2TP205 – "PFA 20:5 (Eicosapentaenoic) (gm)"** **DR1+2TP226 – "PFA 22:6 (Docosahexaenoic) (gm)"**	Summed two-day recall totals of each fatty acid together then divided the value by two to produce an average for each participant. Following this, merged each new value into a "Sum" variable, which accounted for all three FAs in the category.	"Mean Two-Day Total Omega 6 Intakes" "Mean Two-Day Total Omega 3 Intakes"
"Mean Two-Day Total Omega 6 Intakes" **"Mean Two-Day Total Omega 3 Intakes"**	Created a ratio using mean Omega 3 Intakes and mean Omega 6 Intakes.	Calculated Omega 6:3 Fatty Acid Ratio
"Mean Two-Day Total Vitamin A Intakes"	Classified individuals based on WHO recommendations for minimum vitamin A requirements, considering gender-specific requirements. Women who consumed less than 500 mcg a day were considered "Insufficient" while men who consumed less than 625 mcg a day were considered "Insufficient". Sufficiency was coded as "1" and Insufficiency was coded as "2".	"Sufficient/Insufficient Vitamin A Intakes"

MCQ240N – "Age When Lung Cancer First Diagnosed"	A lung cancer diagnosis was considered if a participant had indicated an age for this variable. For the 2017-2018 variables, a response was considered a diagnosis if an individual had filled out a response for 1st cancer diagnosis. Collectively, these responses were aggregated and transformed into a binary variable. Individuals with cancer were considered "1" and individuals without were considered "2".	"Ever Been Diagnosed With Lung Cancer or Malignancy?"
DMEDUC2 – "Education Level – Adults 20+"	Previously an ordinal variable with greater than 5 levels, this variable was recoded to consider individuals with either a high school diploma or greater (1), or less than a high school diploma (2).	"Education Level Highschool Diploma or Greater"
SMD057 – "# cigarettes smoked per day when quit" **SMD070 – "# cigarettes smoked per day now"**	A variable was created to combine both smokers and non-smoker cigarette per day intake. Reports of current smokers and former smokers were aggregated into one, continuous variable.	"Cigarettes Per Day"
SMD030 – "Age started smoking cigarettes regularly"	For current smokers, the variable "Age started smoking cigarettes regularly" was subtracted from the	"Smoking Duration"

SMQ050Q – "How long since quit smoking cigarettes" **&** **SMQ050U – "Unit of measure (day/week/month/year)"** **SMD055 – "Age last smoked cigarettes regularly"** **RIDAGEYR – "Age in Years at Screening"**	variable "Age in Years at Screening" to determine how long an individual had been smoking cigarettes. For non-smokers, the age at which an individual started smoking regularly was subtracted from the age last smoked cigarettes regularly. Together, these values were aggregated into a separate variable and considered "Smoking Duration".	
DR1+2IFDCD – "USDA Food Code"	This variable was used to create four new variables. First, milk drinkers were sought out by looking at the food codes of individuals in the sample. After being identified, the type of milk was categorized from the food codes. Following this, a "Drink Milk?" variable was created with milk drinkers identifying as "1" and non-drinkers identifying as "2". This was carried out with three other variables comparing whole milk drinkers to everyone else, 2% drinkers to everyone else, and whole & 2% drinkers to everyone else.	"Drink Milk?" "Two Percent Consumption vs. Everyone Else" "Whole Milk Consumption vs. Everyone Else" "Two Percent and Whole Milk Consumption vs. Everyone Else"

DBQ197 – "Past 30-day milk product consumption"	Individuals who consumed moderate to frequent amounts of dairy products over the past 30 days were coded as "1" while everyone else was classified as "2".	"Moderate-Frequent Milk Consumption vs. Everyone Else"
BMXWAIST – "Waist Circumference"	Participant waist circumference measurements were organized into sex-specific categories based on whether they were normal or abnormal. 1 = Normal, 2 = abnormal	Abnormal or Normal Waist Circumference

2.3.1. Outcome of Interest

Our primary outcome of interest was lung cancer incidence, the primary variable used to distinguish case and control (see section 2.4, "Exclusion Criteria and Group Selection" below). Lung cancer incidence, while not an explicit variable itself, was derived from participant responses to the "Medical Conditions" questionnaire, a component of the Computer-Assisted Personal Interview (CAPI). The CAPI was given to participants who visited the mobile examination center (MEC) for a medical examination and consists of series of questions deemed too personal for the household interview portion of NHANES.

2.3.2. Demographic Data

Demographic variables of interest included age, gender, number of times seen a healthcare provider over the past year, and education level (Table 1). Age, Waist Circumference, Gender, and Education Level are variables based on questions found in the "Demographic Variables and Sample Weights" data file and are found in each survey cycle. Like lung cancer incidence, these variables are components of the CAPI but administered in the household of the participant. Selection of demographic variables was important for matching cases to controls, exclusion criteria, confounding, and evaluation of descriptive statistics.

2.3.3. Dietary Data

Dietary interview variables of interest included mean two-day vitamin A intakes (totals and subtypes), whether an individual's diet differs from their reported interview intakes, mean essential fatty acid intakes, milk consumption, total dietary

fat, and saturated fat intakes, and usage of multivitamins/vitamin A-containing supplements. Variables associated with vitamin A intakes (and subtypes), essential fatty acid intakes, dietary fat intakes, and milk intakes are found in the "Dietary Interview" data files which contain food and nutrient records based on two 24-hr recalls (Table 1). The first 24-hr recall is conducted in person in a private room, at the MEC. A follow-up recall is then scheduled 3-10 days after the initial visit and conducted via telephone. The specific recall used in this setting is USDA's (United States Department of Agriculture) Automated Multiple Pass Method (AMPM), a recall that is an effective method of reflecting individual nutrient intakes for energy, macronutrients, and micronutrients [27, 28].

In the 03/04 and 05/06 cycles, individual supplement use was found in the dataset "Dietary Supplement Use 30-Day - File 2, Participant's Use of Supplements". This dataset provided codes for each supplement the individual showed the interviewer during the 24-hr recall. The provided codes had to be matched to the specific supplement using the "Dietary Supplement Database - Ingredient Information" data file. This differed from the 07/08 cycles and beyond in that the supplement usage and database information were merged into one file, "Dietary Supplement Use 30 Day - Individual Dietary Supplements" (Table 1). Dietary variables were selected based on our research questions, exclusion criteria when applicable, and the inclusion of potential confounders to our primary outcome.

To ascertain whether participant recall data was like that of their usual intakes, we also were interested in the variable "Compare food consumed yesterday to usual" which allowed participants to address whether the amount of food they reported in

each of their recalls was like their normal intakes. This variable was found in the "Dietary Interview – Total Nutrient Intakes" files for each recall day and was categorized based on whether an individual consumed less than, as much as, or more than their 24-hr recall volume. By doing this, we hoped to increase the likelihood of recall data reflecting long-term dietary intakes/patterns.

2.3.4. Cigarette Smoking Data

Variables included whether an individual smoked cigarettes every day at the time of screening, amount of cigarettes smoked per day, age started smoking cigarettes, amount smoked per day when quit, and computed smoking duration. Importantly, some variables (such as computed smoking duration) had to be computed by combining other variables, as they were only available in some cycles vs others (Table 2). Separately, another variable of interest, "Household Smokers" was found in the "Smoking – Household Smokers" data file.

2.3.5. Healthcare Data

The variable "#times received healthcare over past year" is found in the "Hospital Utilization & Access to Care" data file and is also a component of the CAPI. This variable reports the number of times an individual has seen a healthcare provider over the past year, based on participant-reported frequencies.

2.4. Exclusion Criteria and Case/Control Assignment

Our investigatory population consisted of 115,896 individuals across eight cycles of NHANES surveys (2003-2018), combined according to NHANES analytical guidelines (Johnson et al., 2013). Our reasoning for choosing cycles from these periods was based on the fact that before the 2003-2004 NHANES surveys, only one day of 24-hr dietary recalls was conducted. We thought that relying on dietary estimates from one 24-hr recalls to be truly reflective of long-term nutrient intakes would reduce the validity of our results. Importantly, NHANES recommends weighing survey data using the appropriate sample weights. Given our small sample size, as well as our case-control approach, we decided against weighing our data. While the assigned population weights designated to individuals by NHANES surveyors are meant to account for under-sampling of populations subsets, our case-control approach matched a specific frequency of controls to our case group using well-defined exclusion and matching criteria. If we were to maintain usage of NHANES designated population sample weights, we would lose the matched population proportions derived through our case-control approach. Consequently, this would negate any of the substantial reduction in confounding bias enjoyed through our case-control study design [29].

Variables of interest were identified then combined across cycles (Table 1). Using these variables, a framework for case/control membership was established and applied in two steps (Figure 1). The first step involved the construction of the "Case" group, identified initially as individuals with reported diagnoses of lung cancer per the MEC

questionnaire (Table 1). The second step involved the selection of eligible participants for the control group, followed by a 1:1 matching methodology.

Figure 1 illustrates the assignment of individuals to the "case" group. Broadly, assignment to the "case" group was based on the presence of lung cancer; individuals who reported additional cancer diagnoses were not included as we believed such diagnoses could indicate that there was a chance the lung cancer was not derived organically and metastasized from another bodily location. After this initial selection, individuals who had never smoked cigarettes were removed from our sample, individuals who reported multivitamin/vitamin A-containing supplement usage were removed from the sample, individuals who reported unusual intakes during either of their dietary recall days and individuals with non-response to any variable important for our analyses (Table 1) and research question were excluded from the sample.

The second step, or selection of participants from the control group, relied on an exclusion/selection process that slightly differed from the case group (Figure 1). Starting from the initial pool of the eight NHANES survey cycles, individuals who reported lung cancer or other types of cancer diagnoses were excluded from selection. Following this, individuals with non-response to key variables (Table 1) were excluded from consideration. Individuals were further excluded if they reported smoking at any frequency less than every day of the week. Final exclusions were made if an individual reported multivitamin or vitamin A-containing supplement use, individuals who reported unusual intakes during either of their dietary recall days, and if an individual had seen a healthcare provider less than two times during the previous year. The latter criteria were of particular importance to our outcome, as we posited that individuals who had reported

seeing their healthcare provider for two years or greater over the past year would have less likelihood of an undiagnosed case of lung cancer.

Following the construction of the case group and the selection of eligible participants for the control group, 1:1 matching on key demographic/smoking variables was undertaken on a cycle-to-cycle basis. Starting with individuals with lung cancer in the 03-04 cycle, eligible individuals in the control pool (from the corresponding cycle) were matched on duration smoked, gender, and age. Duration of smoking was selected as the primary matching variable for several reasons. Firstly, lung cancer is a disease that often takes many years to develop and in an estimated 80-90% of cases, is the product of carcinogen exposure to the lungs [30]. As such, we posited that matching on this basis would ensure similar lifetime exposure to smoke-based carcinogens. Further, recent evidence has shown that there is no significant difference between the predictive power of duration smoked (in years) and pack-years, the most used methodology of studying the relationship between cigarette smoking and respiratory disease pathophysiology [31]. This was important to our study design, as the variable indicating the duration of time smoking a reported quantity of cigarettes was only available in the first half of the NHANES cycles which we were analyzing. Thus, it was impossible to compute a "Duration Smoked" variable based on the data available within all eight of our survey cycles. Another important consideration for matching on smoking habits was the amount smoked per day. Given the variability of this metric, the differences in the time of the experimental group from when they last smoked cigarettes and the predictive ability of "duration smoked" compared to other indices of cigarette-based lung cancer risk factors, it was determined that pack years would be a better overall variable to employ. For an

integrated overview of the exclusion criteria, selection framework, and research design, please see figure 1.

2.5. Statistical Approach, Variable Coding, and Covariate Selection

Several statistical approaches were employed to address our research question. Descriptive statistics were gathered on demographic variables, smoking characteristics, and nutrient intakes. Next, paired Samples T-tests were used to analyze differences in means of continuous variables between cases and matched controls. An exact McNemar test was used to assess significant differences in select categorical variable proportions between case and control. Further, conditional logistic regression was used to model the relationship between vitamin A sufficiency/insufficiency and lung cancer incidence in our sample.

Paired Samples T-tests were used to analyze differences in mean total vitamin A and mean total vitamin A subtype intakes between individuals with lung cancer and individuals without lung cancer. As there were two 24-hr recall days that we derived participant nutrient intakes from, the mean total intakes for any nutrient analyzed had to be computed from each recall day. For an overview of this process as it pertains to vitamin A, subtypes of retinoids, and other nutrients of interest please see Table 2. Following this, the variable "Mean total vitamin A intakes" was coded into "sufficient" and "insufficient", based on WHO recommendations for nutritional adequacy. Males and females consuming more than the baseline requirements (625 mcg and 500 mcg, respectively) were coded as "1" while those consuming less than baseline were coded as "2" [32]. Differences in proportions of sufficiency/insufficiency were computed using an exact McNemar Test due to the case-control design of the study.

Next, conditional logistic regression was used to assess the likelihood of lung cancer incidence given vitamin A insufficiency considering confounders of known and exploratory clinical importance. As our study was a case-control design with a control group derived from participants matched on certain characteristics, Cox Regression was manipulated in a manner to produce a conditional logistic regression model [33]. Covariates considered for analysis were "Age Started Smoking Cigarettes Regularly", "Sufficient/Insufficient Vitamin A Intakes", "Household Smokers (Yes/No)", "Education Level", "Drink Milk?", "Calculated Omega 3:6 Fatty Acid Ratio", "Waist Circumference", "Mean Two-Day Dietary Fat Intakes", and "Mean Two-Day Saturated Fat Intakes". In many cases, continuous variables/covariates such as those found in the separate 24-hr recall files had to be combined and averaged to produce singular estimates. Further, independent variables with many levels were recoded to consist of a binary variable in which the responses aggregated many responses into certain brackets. This was both for organization and simplicity of analysis during conditional logistic regression (Table 2)

Following selection and recoding of covariates, we sought to test our hypothesis that vitamin A insufficiency increased the likelihood of lung cancer using conditional logistic regression. Confounders (mentioned above) were analyzed independently and, using an effect size-based approach, were used to fit a model which retained our primary predictor of interest, "Sufficient/Insufficient Vitamin A Intakes". Confounders were retained through both significance and modification of the coefficient of our predictor of interest. Outside of our hypothesis, conditional logistic regression was used to examine confounders from the previous model stage most predictive of lung cancer risk in our

sample, collectively. Univariate analysis of each predictor was conducted, and any predictors with a Wald statistic significant at the p = .25 level were aggregated into a preliminary model. Following this, a model was fit using repeated inclusion/exclusion of covariates and examining changes to both significances of the Wald Statistic and the effect sizes via the beta-coefficient [34].

Before analyses were conducted, assumptions for each statistical approach were checked. Two-tailed significance for events was p = .05, except step-specific significance (described above) during logistic regression model fitting. Means were reported +/- the SD (Standard Deviation) for paired samples t-tests while ORs (Odds Ratios) were reported with 95% confidence intervals. Our analysis was performed using SPSS Version 27 [35].

3. Results

3.1. Demographics and Smoking Characteristics

Table 3 summarizes the demographic and smoking characteristics of individuals in our sample. Our sample consisted of more males (63.7%) males than females (36.3%). Across the sample, there were a greater number of smokers (64.3%) than non-smokers (35.7%). Importantly, around a quarter (28.6%) of individuals with lung cancer currently smoked cigarettes at the time of the survey while most (71.4%) were former smokers. All the matched controls were active, everyday smokers. Individuals with lung cancer were slightly older (M = 66.4, SD = 9.8) than those without lung cancer (M = 62.6, SD = 8.7). Male age (M = 67.2, SD = 10.3) was slightly greater than female age (M = 64.9, SD = 8.7). On average, smokers were older (M = 62.8, SD = 8.9) than non-smokers (M

= 72.83, *SD* = 7.75). On average, the age that those with lung cancer started smoking cigarettes (*M* = 17.4, *SD* = 3.9) was less than the age of those without lung cancer (*M* = 20.0, *SD* = 5.7).

Table 3

Sample-wide differences in total vitamin A intakes & subtypes

Variable	Sample-Wide	Lung Cancer	No Lung Cancer
Mean Two-Day Total Vitamin A Intakes (mcg)	*M* = 535.9, *SD* = 348.0	*M* = 555.4, *SD* = 340.7	*M* = 516.4, *SD* = 357.5
Mean Two-Day Total Retinol Intakes (mcg)	*M* = 378.8, *SD* = 274.7	*M* = 382.5, *SD* = 33.2	*M* = 375.1, *SD* = 44.8
Mean Total Two-Day Carotenoid Intakes (mcg)	*M* = 2101.5, *SD* = 2745.2	*M* = 2295.4, SD = 2518.4	*M* = 1907.6, SD = 2745.2
Mean Two-Day Total Beta Cryptoxanthin Intakes (mcg)	*M* = 130.0, *SD* = 288.2	*M* = 97.8, *SD* = 21.0	*M* = 162.3, *SD* = 54.3
Mean Two-Day Total Alpha Carotene Intakes (mcg)	*M* = 290.3, SD = 538.5	*M* = 312.7, *SD* = 62.4	*M* = 267.9, *SD* = 89.7
Mean Two-Day Total Beta Carotene Intakes (mcg)	*M* = 1681.2, *SD* = 2214.0	*M* = 1884.9, *SD* = 331.1	*M* = 1477.4, *SD* = 301.4

Note: M = "Mean"; SD = Standard Deviation. All reported values are in micrograms (mcg).

3.2. Two-Day Total Vitamin A Intakes, Retinol Intakes, Provitamin A Carotenoid Intakes, and Rates of Sufficiency/Insufficiency Do Not Differ Between Individuals with Lung Cancer and Cancer-Free, Matched Controls

Mean two-day total vitamin A intakes from food were greater in men (M = 555.1, SD = 356.8) then in women (M = 503.0, SD = 334.6). Mean two-day total vitamin A intakes from food were greater in non-smokers (M = 614.3, SD = 354.5) compared to smokers (M = 492.4, SD = 339.3). Sample-wide differences in total vitamin A intakes, as well as intakes of vitamin A subtypes, can be found in Table 3. A paired samples t-test showed no significant difference in Mean Two-Day Total Vitamin A Intakes between those diagnosed with lung cancer **(Table 4)**. Additional paired samples t-tests showed no significant differences in mean intakes of retinol, total carotenoids, or subtypes of provitamin A carotenoids (Table 4). Next, individual Mean Two-Day Total Vitamin A Intakes were classified into a binary "Sufficient or Insufficient" variable as outlined in Table 2. An exact McNemar Test determined that there was no significant difference in the proportion of sufficient/insufficient vitamin A intakes between case and control (1, N = 49), p = .84.

Table 4

Mean Differences in Total Vitamin A Intakes and Vitamin A Subtypes Between Matched Case & Control

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair 1	Mean Two-Day Total Vitamin A Intakes (mcg)	39.0	547.1	78.16	-118.15	196.15	.50	48	.62
Pair 2	Mean Two-Day Total Retinol Intakes (mcg)	7.4	419.9	59.98	-113.25	127.95	.12	48	.90
Pair 3	Mean Total Two-Day Carotenoid Intakes (mcg)	387.8	3815.6	545.08	-708.15	1483.78	.71	48	.48
Pair 4	Mean Two-Day Total Beta Cryptoxanthin Intakes (mcg)	-64.5	412.8	58.98	-183.08	54.08	-1.09	48	.28
Pair 5	Mean Two-Day Total Alpha Carotene Intakes (mcg)	44.8	792.7	113.24	-182.88	272.49	.40	48	.69

Pair 6	Mean Two-Day Total Beta Carotene Intakes (mcg)	407.5	3195.7	456.53	-510.42	1325.40	.89	48	.38

Note: t = "T-value"; df = "Degrees of Freedom". All reported values are in micrograms (mcg).

* $p < 0.05$

3.3. Individuals With Lung Cancer Are More Likely to Consume Minimally Sufficient Intakes of Vitamin A Then Individuals Without Lung Cancer, Adjusting for Education Level, Waist Circumference, Age Started Smoking, Milk and EFA Intakes

We used conditional logistic regression to test our hypothesis that vitamin A insufficiency increases the likelihood of lung cancer. Before adjustment for potential confounding, individuals with lung cancer were less likely to consume sufficient intakes of vitamin A (OR = 0.85, CI = 0.38 – 1.89), but this association was not significant. This was not surprising, as an exact McNemar Test determined that there was no statistically significant difference in the proportion of individuals with insufficient mean total Vitamin A intakes between case and controls (1, N = 49), p = .84. Adjusting for education, age of smoking initiation, Omega 6:3 Essential Fatty Acid Ratio, waist circumference, and milk consumption individuals with lung cancer were nearly 5.5 times more likely to consume minimally-sufficient amounts of vitamin A, compared to individuals without lung cancer **(Table 5)**

3.4. Age Started Smoking Regularly, Household Smokers, Omega 6:3 Fatty Acid Ratio and Milk Consumption Are Predictors of Lung Cancer Risk in a Multivariate Conditional Logistic Regression Model

Using conditional logistic regression, we sought to look at all predictors to assess which both univariately and multivariately best-predicted lung cancer incidence in our sample. Selected covariates that were univariately predictive of lung cancer at the preliminary *p* = .25 level were “Age started smoking cigarettes regularly”, “Does anyone smoke inside home?”, “Education Level”, “Milk Drinker?”, and “Calculated Omega 6:3 Fatty Acid Ratio” **(Table 6).** Next, a new model was fit as described in section 2.5, “*Methods”*.

After refitting our initial model several times to account for the effects of confounding covariates on coefficient values, variables with the greatest multivariate impact on lung cancer likelihood were “Age started smoking cigarettes regularly”, “Does anyone smoke inside home?”, “Calculated Omega 6:3 Fatty Acid Ratio”, and “Drink Milk?” **(Table 7).**

Table 5

Model incorporating vitamin A sufficiency/insufficiency adjusted for education level, waist circumference, age started smoking, milk drinking, and omega 6:3 ratio

	B	SE	Wald	df	Sig.	Exp(B)	95.0% CI for Exp(B)	
							Lower	Upper
Sufficient/Insufficient Vitamin A Intakes	1.70	.97	3.06	1	.080	5.49	.82	36.96
Age started smoking cigarettes regularly (years)	-.53	.22	6.05	1	.014	.59	.38	.90
Education Level	-1.59	1.08	2.18	1	.14	.20	.03	1.69
Waist Circumference	-1.59	1.08	2.18	1	.14	.20	.03	1.69
Drink Milk?	1.20	.86	1.92	1	.17	3.31	.61	18.05
[a] **Calculated Omega 6:3 Fatty Acid Ratio**	-.27	.18	2.16	1	.14	.77	.54	1.09

Note. B = "Beta Coefficient"; SE = "Standard Error" associated with beta coefficient; Wald = "Wald Chi-Square Value"; df = "Degrees of Freedom"; Sig. = Significance of Wald Chi-Square Statistic; Exp(B) = "Odds Ratio".

[a] Covariates without units of measurement next to their name are categorical, except for "Calculated Omega 6:3 Fatty Acid Ratio". This variable was derived by computing a ratio from Omega 3 fatty acid (gm) and Omega 6 fatty acid (gm).

* $p < .025$.

Table 6

Univariate analysis of included covariates

	B	SE	Wald	df	Sig.	Exp(B)	95.0% CI for Exp(B)	
							Lower	Upper
Sufficient/Insufficient Vitamin A Intakes	-.17	.41	.16	1	.68	.85	.53	2.64
Age started smoking cigarettes regularly (years)	-.34	.11	9.06	1	.003*	.71	.57	.89
Does anyone smoke inside home?	.99	.44	5.10	1	.02*	2.71	1.14	6.46
Education level	.56	.44	1.60	1	.21*	1.75	.73	4.17
[a] **Calculated Omega 6:3 Fatty Acid Ratio**	-.07	.05	1.86	1	.17*	.93	.84	1.03
Waist Circumference	-.36	.49	.52	1	.47	.70	.27	1.84
Drink Milk?	.87	.42	4.2	1	.04*	2.38	1.04	5.43
Mean Two-Day Dietary Fat Intakes (gm)	-.01	.01	.91	1	.34	.99	.98	1.01
Meant Two-Day Saturated Fat Intakes (gm)	-.01	.02	.29	1	.59	.99	.96	.102

Note. B = “Beta Coefficient”; SE = “Standard Error” associated with beta coefficient; Wald = “Wald Chi-Square Value”; df = “Degrees of Freedom”; Sig. = Significance of Wald Chi-Square Statistic; Exp(B) = “Odds Ratio”.

[a] Covariates without units of measurement next to their name are categorical, except for "Calculated Omega 6:3 Fatty Acid Ratio". This variable was derived by computing a ratio from Omega 3 fatty acid (gm) and Omega 6 fatty acid (gm).

* $p < .025$.

Table 7

Sample-specific predictive model for lung cancer incidence

	B	SE	Wald	df	Sig.	Exp(B)	95.0% CI for Exp(B)	
							Lower	Upper
Age started smoking cigarettes regularly (years)	-.41	.16	6.90	1	.01	.67	.49	.90
Household Smokers	1.30	.58	4.98	1	.03	3.66	1.17	11.46
[a] **Calculated Omega 6:3 Fatty Acid Ratio**	-.14	.09	2.31	1	.13	.87	.73	1.04
Drink Milk?	1.20	.64	3.50	1	.06	3.32	.94	11.65

Note. B = "Beta Coefficient"; SE = "Standard Error" associated with beta coefficient; Wald = "Wald Chi-Square Value"; df = "Degrees of Freedom"; Sig. = Significance of Wald Chi-Square Statistic; Exp(B) = "Odds Ratio".

[a] Covariates without units of measurement next to their name are categorical, except for "Calculated Omega 6:3 Fatty Acid Ratio". This variable was derived by computing a ratio from Omega 3 fatty acid (gm) and Omega 6 fatty acid (gm).

* $p < .025$.

3.5. Milk Consumption Increases Odds of Lung Cancer Incidence

Individuals who consumed any type of milk during one or both of their 24-hr recall days were nearly 2.5 times as likely to have lung cancer compared to those who reported no milk consumption (Table 6). The covariate also contributed a substantial amount of effect modification to both our vitamin A-dependent predictive model (Table 6). Given these results, we descriptively analyzed smoking, demographic, and dietary characteristics of milk drinkers in our sample to better understand any connections between our covariates and the outcome. Importantly, our research design was a case-control study using pairwise matching; consequently, we were limited to comparisons of descriptive statistics between the groups and were not able to determine statistical significance. Caution should be taken when drawing conclusions from these results.

Table 8 outlines comparative frequencies of smoking, demographic, and dietary characteristics between milk drinkers and non-milk drinkers. Notably, a greater percentage of non-milk drinkers were smokers (75.5%) compared to those who did drink milk (53.1%). Further, vitamin A insufficiency was greater in non-milk drinkers (71.4%) than milk drinkers (59.2%). Across continuous variables, characteristics of milk drinkers did not substantially differ from non-milk drinkers apart from mean two-day total vitamin A intakes and mean two-day total retinol intakes. Milk drinkers consumed more total vitamin A from food compared to non-milk drinkers, and more retinol from food compared to non-milk drinkers **(Table 9)**

We further validated this result with data from the “Diet Behavior & Nutrition” questionnaire portion of the NHANES surveys. Using the variable “Past 30-day milk product consumption”, we created a variable that combined both this and the “Milk

Drinker?" variable. We called this variable "True Milk Drinkers"; individuals were considered a "True Milk Drinker" if they reported milk consumption from our original variable while additionally reporting either moderate to frequent milk consumption in the "Past 30-day milk product consumption" category. After this variable combination and recoding, we found that individuals who were both milk drinkers and moderate to frequent dairy consumers were nearly three times as likely to have lung cancer, compared to those who did not fall into either category (OR = 2.83, CI = 1.12 – 7.19).

Table 8

Comparisons of demographic, smoking, and dietary characteristics between those who do and do not drink milk

	Non-Milk Drinkers			Milk Drinkers	
	N	Mean	Std. Deviation	Mean	Std. Deviation
Age started smoking cigarettes regularly (years)	49	18.9	5.4	18.5	4.8
Mean Two-Day Total Vitamin A Intakes (mcg)	49	476.4	374.5	595.5	311.7
Mean Two-Day Total Beta Cryptoxanthin Intakes (mcg)	49	138.1	278.6	122.0	300.3
Mean Two-Day Total Alpha Carotene Intakes (mcg)	49	298.6	644.7	281.9	412.5
Mean Two-Day Total Beta Carotene Intakes (mcg)	49	1623.4	2110.3	1738.9	2333.5
Mean Two-Day Total Retinol Intakes (mcg)	49	323.2	299.3	434.4	238.0
Mean Total Two-Day Carotenoid Intakes (mcg)	49	2060.2	2714.4	2142.8	2565.8
Calculated Omega 6:3 Fatty Acid Ratio	49	10.6	6.1	9.7	3.4
Mean Two-Day Total Fat Intakes (gm)	49	74.2	44.4	73.4	29.1

Mean Two-Day Saturated Fat Intakes (gm)	49	23.6	12.9	24.8	11.6

Table 9

Frequencies of categorical demographic, smoking, and dietary variables in milk drinkers and non-milk drinkers

Variable		**Non-Milk Drinkers**		**Milk Drinkers**	
	Sub-category	N	%	N	%
Do you now smoke cigarettes	Yes	37	75.5%	26	53.1%
	No	12	24.5%	23	46.9%
Sufficient/Insufficient Vitamin A Intakes	Sufficient	14	28.6%	20	40.8%
	Insufficient	35	71.4%	29	59.2%
Does anyone smoke inside the home?	Yes	20	40.8%	22	44.9%
	No	29	59.2%	27	55.1%
Education Level		30	61.2%	34	69.4%

	At least high school				
	Lower than high school	19	38.8%	15	30.6%
Gender	Male	31	63.3%	31	63.3%
	Female	18	36.7%	18	36.7%

4. Discussion

In our case-control study examining vitamin A sufficiency/insufficiency and its impact on the likelihood of lung cancer incidence, we found that milk drinkers were nearly 2.5 times as likely to have been diagnosed with lung cancer, compared to those who reported no milk consumption. Contrary to our hypothesis, we found an increased association of lung cancer in individuals consuming minimally sufficient amounts of vitamin A, following adjustment for age of smoking initiation, omega 3:6 ratio, waist circumference, milk consumption, and education level. Between case and control, no significant differences in average two-day intakes of total vitamin A or subtypes were observed.

The increased likelihood of lung cancer with milk consumption agrees with epidemiological evidence showing increased risk of lung cancer amongst current and former smokers who consume whole-fat dairy products [36-39]. Nutrients in milk studied for their associations with lung cancer include dietary fat, vitamin D, retinol, and calcium; as retinol was a variable of interest in our design, we showed that differences in mean two-day intakes of retinol were not significantly different between case and control (Table 4). Compared to non-milk drinkers, milk drinkers consumed 25.0% more total vitamin A from the diet and 34.4% more total retinol from the diet. Given these findings, it may be plausible that our finding of vitamin A sufficiency's impact on lung cancer likelihood is related to retinol in milk consumption. Due to the pairwise matching employed with our case-control design, however, we were unable to assess whether these differences were statistically significant. Further, we did not collect data surrounding the contributions of nutrients such as retinol from milk, to overall intakes.

Supplemental dosages of retinol, in combination with beta-carotene, show positive associations with lung cancer risk; in the Beta-Carotene and Retinol Efficacy Trial (CARET), a large interventional trial used to reduce the risk of lung cancer through supplementation of retinoids, administration of a combination of high-dose beta-carotene and retinol significantly increased the likelihood of lung cancer [40]. It was hypothesized that the supraphysiological dosage of carotenoids and vitamin A came in contact with oxidatively damaging components of cigarette smoke leading to the formation of biologically inert/dangerous retinoid byproducts, as well as antagonists to vitamin receptors. The administered dosage of retinol alone, however, was greater than 10 times the intakes of the case, control, milk drinkers, and non-milk drinkers in our sample. Consequently, the impact of milk consumption on the likelihood of lung cancer observed in our sample may be a product of other bioactive components of milk in tandem with or independent of retinol, even in those who were everyday smokers.

Outside retinol, cow's milk contains macronutrients and micronutrients known to impact lung cancer risk; such nutrients include fat, calcium and vitamin D. Saturated fat has been shown to increase the likelihood of lung cancer; in our sample, milk drinkers predominately consumed 2% or whole milk, suggesting that intakes of saturated fat may be greater in this subsample [41]. Intakes of total dietary fat and saturated fat from food, however, were not independently associated with any significant increase or decrease in the likelihood of lung cancer incidence in our study, nor were they realized as contributing confounders to our final adjusted model. Further, no significant difference between mean total dietary fat and mean total saturated fat was found between case and control, and minimal differences in intakes between milk drinkers and non-milk drinkers were observed; the latter observation, however,

was not able to be tested for significance and caution should be given towards interpretation of this result.

We did not calculate intakes of dietary vitamin D and calcium by each participant; as such, we cannot rule out residual confounding or influence by either of these micronutrients on lung cancer risk. High intakes of dietary vitamin D, but not circulating serum levels of vitamin D, were shown to reduce the risk of lung cancer in a meta-analysis that pooled 16 different studies [42]. Vitamin D has been shown to exhibit anticarcinogenic actions through its inhibition of signaling pathways including those that cause mutations in Wnt /β-catenin, epidermal growth factor receptor, and vascular endothelial growth factor; collectively these signaling pathways regulate cell proliferation, differentiation, genetic stability, apoptosis, and angiogenesis, and play a role in lung tumorigenesis. Further, vitamin D promotes cell cycle arrest via the cyclin-dependent kinase inhibitors p21 and p27; interestingly, mutations in these specific proteins are associated with lung cancer subtypes typical of non-smokers, suggesting vitamin D adequacy may play a role in preventing lung cancer in never-smokers [43, 44].

Calcium is another micronutrient that has been shown to have implications in lung carcinogenesis. Amongst early-stage lung cancer patients, low pre-diagnostic dietary calcium has been associated with poor survival while intakes of calcium, magnesium, and phosphorous are inversely associated with lung cancer in non-smokers [45, 46]. Insufficient calcium intakes, particularly over long periods, can lead to activation of pathways implicated in tumor growth and metastasis such as IL-6 (Interleukin-6), Vascular Endothelial Growth Factor, and Macrophage Colony-stimulating Factor [46]. Further, calcium insufficiency and subsequent increases in bone resorption increase the likelihood of tumor metastasis to the

bone; as an estimated 40% of metastatic lung cancer results in bone tumors, calcium sufficiency may be of particular importance to individuals with lung cancer, especially if they smoke cigarettes [46, 47]. In contrast, intakes of calcium have also been shown to be associated with an increased risk of lung cancer, an effect which was made stronger when accounting for the effects of zinc and iron in the model; these associations were stronger in current smokers than former smokers [48]. Apart from the latter study, however, these results are not in line with our finding that milk consumption increased the likelihood of lung cancer incidence, especially since milk is a rich source of both calcium and vitamin D. It should be noted, however, that since we did not calculate intakes of each micronutrient, we cannot conclude that these nutrients did not impact lung cancer likelihood in our sample.

While milk drinking was strongly and independently associated with an increased incidence of lung cancer it was also a confounder in our adjusted model which retained our primary independent variable of interest, sufficient/insufficient mean two-day vitamin A intakes, to investigate its impact on lung cancer incidence. As previously discussed, while milk drinkers consumed 25.0% more total vitamin A from the diet and 34.4% more total retinol from the diet compared to non-milk drinkers, we were not able to determine whether such differences were statistically significant due to the design of our study. Further, the differences between intakes were not observably large in either category; as supraphysiological intakes of retinol and beta-carotene increased risk for lung cancer in smokers, an appreciably large intake of total vitamin A may be necessary to increase the risk of lung cancer, especially in the presence of cigarette smoke [40].

As we were not able to rule out confounding effects from age of smoking initiation and the variable was independently and multivariately associated with an increased likelihood of

lung cancer incidence, the possibility of a diet-cigarette interaction should not be ruled out. Earlier smoking initiation, especially of cigarettes, increases the risk of smoking-related morbidities and all-cause mortality [49]. Cigarette smoke has been shown to impact levels of vitamin A in the lung and increase the requirements for dietary vitamin E in humans [13, 50]. Given the remarkably low levels of mean vitamin A intakes across both case and control, it may be that age of smoking initiation combined with poor nutrient intakes across the lifespan has a collectively carcinogenic effect. While we matched individuals on smoking duration and eliminated a likely source of variability in our findings, the confounding conferred by age of smoking initiation to our model and its interplay with lifetime intakes of certain nutrients warrants further investigation.

We were not able to rule out confounding effects of waist circumference in our model; generally, increased waist circumference has been associated with an increased risk of a spectrum of chronic diseases, such as Coronary Artery Disease, Type II Diabetes, and Hypertension [51-53]. Interestingly, increased waist circumference has also been associated with the risk of several forms of cancer, including lung cancer [54]. While we did not calculate the caloric intake of participants in our study, abnormal waist circumference could be an indicator of overconsumption of calories, and consequently, a nutrient-poor diet [55]. Thus, the impact of this covariate in our adjusted model on lung cancer risk could signify poor food choices, poor lifestyle choices, low nutrient intakes, and subsequent increased risk of chronic disease such as cancer. As there were no significant differences between mean two-day total vitamin A or vitamin A subtypes intakes between case and control, this result would make sense, particularly as only 20.4% of individuals in our sample consumed greater than or equal to gender-specific RDAs for vitamin A. As individuals in our case group had a proportionally

lower amount of individuals with abnormal waist circumferences, however, such a conclusion is not consistent with our findings. More research into energy intakes, waist circumference, abdominal obesity, micronutrients, and their collaborative impact on lung cancer risk is necessitated.

We could not rule out confounding by the calculated Omega 6:3 essential fatty acid ratio in our adjusted model. Univariately, as the calculated Omega 6:3 essential fatty acid ratio increased in our sample, lung cancer was less likely. Omega 3 fatty acids can modulate inflammatory processes through antagonizing and inhibiting arachidonic acid-based inflammatory pathways [56]. Further, Docosahexaenoic acid (DHA) and Eicosapentaenoic acid (EPA), which are synthesized from alpha-linolenic acid, have been shown to inhibit metastatic and growth pathways associated with lung cancer [57-59]. As dietary antioxidants are thought to have an anti-inflammatory effect and be protective against lung cancer in both smokers and non-smokers, a synergistic protective effect between anti-inflammatory fatty acids and pro-vitamin A carotenoids would make sense [18, 19]. A simultaneous reduction in lung cancer risk might be expected in such a scenario, but this was not illustrated by our findings.

To our knowledge, our case-control approach was the first of its type to investigate vitamin A insufficiency and its impact on lung cancer incidence, using pairwise matching of case and control based on the duration of cigarette smoking. The strength of this design lies in this matching criteria, as smoking duration is an effective predictor of lung [31, 60, 61]. Further, matching individuals in this manner allowed us to rule out a significant source of confounding concerning diet and lung cancer incidence. With this strength comes a few important weaknesses to our design. While we were able to rule out confounding by smoking

duration through our research design, we could not rule out residual confounding by smoking intensity. We did not include this variable as part of our exclusion criteria due to the proportionally greater number of former smokers in our "case" group, compared to current smokers. We believed that the introduction of such a variable could introduce a source of bias into our results, as the importance of smoking intensity would likely diminish as the time in years since quitting smoking in former smokers increased. Importantly, however, it has been shown that the duration of smoking in years is as good a predictor of lung carcinogenesis as both pack-year calculations and smoking intensity [31].

Another weakness in our research design was that it relied on 24-hr recalls across two days to establish estimated nutrient intakes from foods. Thus, we could not rule out the potential for underreporting of foods when these recalls were conducted, especially as the second of the two recalls were conducted over the telephone. Further, we relied on participant's accounts of whether their dietary recall reports were consistent with the usual amount/type of food that they consume. It should be noted, however, that we did control for supplemental intakes of vitamin A in our design through the exclusion of individuals who reported intakes of multivitamins or retinol-containing supplements. Thus, while potential underreporting of total dietary intakes of vitamin A could have occurred, the likelihood of reported intakes vastly exceeding reported amounts is unlikely. Further, the 24-hr recall methodology used during NHANES data collection has been shown to accurately estimate macronutrients, micronutrients, and energy from the diet [27, 28]. Other weaknesses in our study include the small sample size, the retrospective design, overall poor reported intakes of vitamin A across our group, and reliance on participant-reported cases of lung cancer rather than clinically validated cases through electronic medical records or similar resources

In conclusion, the results of our case-control, exploratory study show a positive association between milk consumption and lung cancer, in current and former smokers. While sufficient vitamin A intakes were less likely in individuals with lung cancer, the crude OR was not significant. After adjustment for clinically/statistically important confounders, the association was reversed. As we could not rule out the confounding effects of smoking, milk consumption, and other important confounders and as total vitamin A was not significantly different between case and control, it may be that the effect conferred by vitamin A sufficiency on lung cancer risk is attributed by other nutritional components which were unmeasured in our study. As such, the increased vitamin A intake and their effects on lung cancer risk may be a surrogate marker for increased nutrients/energy elsewhere. Further, the relatively low levels of reported vitamin A intakes in our study compared to gender-specific RDAs may mean that the effects of vitamin A on lung cancer were too low across the sample. This may mean that a meaningful effect of sufficient vitamin A intakes on lung cancer likelihood may not have been possible to observe through our research design. Overall, our exploratory study is hypothesis-generating, and our major finding calls for further investigation into the role of nutrients and other bioactive components of milk (and other vitamin A-rich foods) which modify lung cancer risk, either with or without vitamin A.

5. References for Body of Thesis

1. Das BC, Thapa P, Karki R, Das S, Mahapatra S, Liu TC, Torregroza I, Wallace DP, Kambhampati S, Van Veldhuizen P, Verma A, Ray SK, Evans T. Retinoic acid signaling pathways in development and diseases. Bioorg Med Chem. 2014 Jan 15;22(2):673-83. doi: 10.1016/j.bmc.2013.11.025. Epub 2013 Nov 22. PMID: 24393720; PMCID: PMC4447240.

2. WHO. Global prevalence of vitamin A deficiency in populations at risk 1995–2005. WHO Global Database on Vitamin A Deficiency. Geneva, World Health Organization, 2009.

3. WHO, Micronutrient Deficiencies.WHO Nutrition Home Page. Geneva, World Health Organization, 2020.

4. Karim T, Muhit M, Khandaker G. Interventions to prevent respiratory diseases - Nutrition and the developing world. Paediatr Respir Rev. 2017 Mar;22:31-37. doi: 10.1016/j.prrv.2016.09.003. Epub 2016 Sep 28. PMID: 27793738.

5. Timoneda J, Rodríguez-Fernández L, Zaragozá R, Marín MP, Cabezuelo MT, Torres L, Viña JR, Barber T. Vitamin A Deficiency and the Lung. Nutrients. 2018 Aug 21;10(9):1132. doi: 10.3390/nu10091132. PMID: 30134568; PMCID: PMC6164133.

6. Baybutt RC, Hu L, Molteni A. Vitamin A deficiency injures lung and liver parenchyma and impairs function of rat type II pneumocytes. J Nutr. 2000 May;130(5):1159-65. doi: 10.1093/jn/130.5.1159. PMID: 10801913.

7. Luo ZX, Liu EM, Luo J, Li FR, Li SB, Zeng FQ, Qu P, Fu Z, Li TY. Vitamin A deficiency and wheezing. World J Pediatr. 2010 Feb;6(1):81-4. doi: 10.1007/s12519-010-0012-7. Epub 2010 Feb 9. PMID: 20143217.

8. Chen F, Marquez H, Kim YK, Qian J, Shao F, Fine A, Cruikshank WW, Quadro L, Cardoso WV. Prenatal retinoid deficiency leads to airway hyperresponsiveness in adult mice. J Clin Invest. 2014 Feb;124(2):801-11. doi: 10.1172/JCI70291. Epub 2014 Jan 9. PMID: 24401276; PMCID: PMC3904614.

9. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. CA Cancer J Clin. 2019 Jan;69(1):7-34. doi: 10.3322/caac.21551. Epub 2019 Jan 8. PMID: 30620402.

10. Pesch B, Kendzia B, Gustavsson P, Jöckel KH, Johnen G, Pohlabeln H, Olsson A, Ahrens W, Gross IM, Brüske I, Wichmann HE, Merletti F, Richiardi L, Simonato L, Fortes C, Siemiatycki J, Parent ME, Consonni D, Landi MT, Caporaso N, Zaridze D, Cassidy A, Szeszenia-Dabrowska N, Rudnai P, Lissowska J, Stücker I, Fabianova E, Dumitru RS, Bencko V, Foretova L, Janout V, Rudin CM, Brennan P, Boffetta P, Straif K, Brüning T. Cigarette smoking and lung cancer--relative risk estimates for the major histological types from a pooled analysis of case-control studies. Int J Cancer. 2012 Sep 1;131(5):1210-9. doi: 10.1002/ijc.27339. Epub 2011 Dec 14. PMID: 22052329; PMCID: PMC3296911.

11. Goldkorn T, Filosto S, Chung S. Lung injury and lung cancer caused by cigarette smoke-induced oxidative stress: Molecular mechanisms and therapeutic opportunities involving the ceramide-generating machinery and epidermal growth factor receptor. Antioxid

Redox Signal. 2014 Nov 20;21(15):2149-74. doi: 10.1089/ars.2013.5469. Epub 2014 Jul 1. PMID: 24684526; PMCID: PMC4215561.

12. Xue Y, Harris E, Wang W, Baybutt RC. Vitamin A depletion induced by cigarette smoke is associated with an increase in lung cancer-related markers in rats. J Biomed Sci. 2015 Oct 14;22:84. doi: 10.1186/s12929-015-0189-0. PMID: 26462767; PMCID: PMC4605095.

13. Li T, Molteni A, Latkovich P, Castellani W, Baybutt RC. Vitamin A depletion induced by cigarette smoke is associated with the development of emphysema in rats. J Nutr. 2003 Aug;133(8):2629-34. doi: 10.1093/jn/133.8.2629. PMID: 12888649.

14. Sung WJ, Kim H, Park KK. The biological role of epithelial-mesenchymal transition in lung cancer (Review). Oncol Rep. 2016 Sep;36(3):1199-206. doi: 10.3892/or.2016.4964. Epub 2016 Jul 21. PMID: 27460444.

15. Koren A, Motaln H, Cufer T. Lung cancer stem cells: a biological and clinical perspective. Cell Oncol (Dordr). 2013 Jul;36(4):265-75. doi: 10.1007/s13402-013-0141-9. Epub 2013 Jul 31. PMID: 23900738.

16. De Craene B, Berx G. Regulatory networks defining EMT during cancer initiation and progression. Nat Rev Cancer. 2013 Feb;13(2):97-110. doi: 10.1038/nrc3447. PMID: 23344542.

17. Esteban-Pretel G, Marín MP, Renau-Piqueras J, Barber T, Timoneda J. Vitamin A deficiency alters rat lung alveolar basement membrane: reversibility by retinoic acid. J Nutr Biochem. 2010 Mar;21(3):227-36. doi: 10.1016/j.jnutbio.2008.12.007. Epub 2009 Mar 9. PMID: 19269151.

18. Vieira AR, Abar L, Vingeliene S, Chan DS, Aune D, Navarro-Rosenblatt D, Stevens C, Greenwood D, Norat T. Fruits, vegetables and lung cancer risk: a systematic review and meta-analysis. Ann Oncol. 2016 Jan;27(1):81-96. doi: 10.1093/annonc/mdv381. Epub 2015 Sep 14. PMID: 26371287.

19. Shareck M, Rousseau MC, Koushik A, Siemiatycki J, Parent ME. Inverse Association between Dietary Intake of Selected Carotenoids and Vitamin C and Risk of Lung Cancer. Front Oncol. 2017 Feb 28;7:23. doi: 10.3389/fonc.2017.00023. PMID: 28293540; PMCID: PMC5328985.

20. Yang T, Wang C, Li S, Guo XF, Li D. Dietary intakes of fruits and vegetables and lung cancer risk in participants with different smoking status: a meta-analysis of prospective cohort studies. Asia Pac J Clin Nutr. 2019;28(4):770-782. doi: 10.6133/apjcn.201912_28(4).0014. PMID: 31826375.

21. Yu N, Su X, Wang Z, Dai B, Kang J. Association of Dietary Vitamin A and β-Carotene Intake with the Risk of Lung Cancer: A Meta-Analysis of 19 Publications. Nutrients. 2015 Nov 11;7(11):9309-24. doi: 10.3390/nu7115463. PMID: 26569298; PMCID: PMC4663591.

22. McClure JB, Divine G, Alexander G, Tolsma D, Rolnick SJ, Stopponi M, Richards J, Johnson CC. A comparison of smokers' and nonsmokers' fruit and vegetable intake and relevant psychosocial factors. Behav Med. 2009 Spring;35(1):14-22. doi: 10.3200/BMED.35.1.14-22. PMID: 19297300; PMCID: PMC2687811.

23. Zhai T, Li S, Hu W, Li D, Leng S. Potential Micronutrients and Phytochemicals against the Pathogenesis of Chronic Obstructive Pulmonary Disease and Lung Cancer. Nutrients.

2018 Jun 25;10(7):813. doi: 10.3390/nu10070813. PMID: 29941777; PMCID: PMC6073117.

24. Centers for Disease Control and Prevention. (2020). National Center for Health Statistics: NHANES Questionnaires, Datasets, and Related Documentation. Retrieved from https://wwwn.cdc.gov/nchs/nhanes/Default.aspx

25. Chen TC, Clark J, Riddles MK, Mohadjer LK, Fakhouri THI. National Health and Nutrition Examination Survey, 2015−2018: Sample design and estimation procedures. National Center for Health Statistics. Vital Health Stat 2(184). 2020.

26. Tenny S, Kerndt CC, Hoffman MR. Case Control Studies. 2020 Jul 10. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2021 Jan–. PMID: 28846237.

27. Moshfegh AJ, Rhodes DG, Baer DJ, Murayi T, Clemens JC, Rumpler WV, Paul DR, Sebastian RS, Kuczynski KJ, Ingwersen LA, Staples RC, Cleveland LE. The US Department of Agriculture Automated Multiple-Pass Method reduces bias in the collection of energy intakes. Am J Clin Nutr. 2008 Aug;88(2):324-32. doi: 10.1093/ajcn/88.2.324. PMID: 18689367.

28. Blanton CA, Moshfegh AJ, Baer DJ, Kretsch MJ. The USDA Automated Multiple-Pass Method accurately estimates group total energy and nutrient intake. J Nutr. 2006 Oct;136(10):2594-9. doi: 10.1093/jn/136.10.2594. PMID: 16988132.

29. Li Y, Graubard BI, DiGaetano R. Weighting methods for population-based case-control studies with complex sampling. Journal of the Royal Statistical Society: Series C (Applied Statistics). 2011;60:165-185. https://doi.org/10.1111/j.1467-9876.2010.00731.x

30. Avci N, Hayar M, Altmisdortoglu O, Tanriverdi O, Deligonul A, Ordu C, Evrensel T. Smoking habits are an independent prognostic factor in patients with lung cancer. Clin Respir J. 2017 Sep;11(5):579-584. doi: 10.1111/crj.12386. Epub 2015 Oct 7. PMID: 26365261.

31. Ogawa K, Koh Y, Kaneda H, Izumi M, Matsumoto Y, Sawa K, Fukui M, Taniguchi Y, Yoshimoto N, Tamiya A, Ando M, Kubo A, Isa SI, Saka H, Matsumura A, Kawaguchi T. Can smoking duration alone replace pack-years to predict the risk of smoking-related oncogenic mutations in non-small cell lung cancer? A cross-sectional study in Japan. BMJ Open. 2020 Sep 9;10(9):e035615. doi: 10.1136/bmjopen-2019-035615. PMID: 32907893; PMCID: PMC7482473.

32. Institute of Medicine (US) Panel on Micronutrients. Dietary Reference Intakes for Vitamin A, Vitamin K, Arsenic, Boron, Chromium, Copper, Iodine, Iron, Manganese, Molybdenum, Nickel, Silicon, Vanadium, and Zinc. Washington (DC): National Academies Press (US); 2001. PMID: 25057538.

33. Conditional Logistic Regression Using COXREG. IBM Support; 2020 April [cited 2021 June 17]. Available from: https://www.ibm.com/support/pages/conditional-logistic-regression-using-coxreg

34. Zhang Z. Model building strategy for logistic regression: purposeful selection. Ann Transl Med. 2016 Mar;4(6):111. doi: 10.21037/atm.2016.02.15. PMID: 27127764; PMCID: PMC4828741.

35. IBM SPSS Statistics for Windows, version 27 (IBM Corp., Armonk, N.Y.,

36. Axelsson G, Liljeqvist T, Andersson L, Bergman B, Rylander R. Dietary factors and lung cancer among men in west Sweden. Int J Epidemiol. 1996 Feb;25(1):32-9. doi: 10.1093/ije/25.1.32. PMID: 8666501.

37. Mettlin C. Milk drinking, other beverage habits, and lung cancer risk. Int J Cancer. 1989 Apr 15;43(4):608-12. doi: 10.1002/ijc.2910430412. PMID: 2703270.

38. Mettlin CJ, Schoenfeld ER, Natarajan N. Patterns of milk consumption and risk of cancer. Nutr Cancer. 1990;13(1-2):89-99. doi: 10.1080/01635589009514049. PMID: 2300498.

39. Axelsson G, Rylander R. Diet as risk for lung cancer: a Swedish case-control study. Nutr Cancer. 2002;44(2):145-51. doi: 10.1207/S15327914NC4402_04. PMID: 12734060.

40. Omenn GS, Goodman GE, Thornquist MD, Balmes J, Cullen MR, Glass A, Keogh JP, Meyskens FL, Valanis B, Williams JH, Barnhart S, Hammar S. Effects of a combination of beta carotene and vitamin A on lung cancer and cardiovascular disease. N Engl J Med. 1996 May 2;334(18):1150-5. doi: 10.1056/NEJM199605023341802. PMID: 8602180.

41. Yang JJ, Yu D, Takata Y, Smith-Warner SA, Blot W, White E, Robien K, Park Y, Xiang YB, Sinha R, Lazovich D, Stampfer M, Tumino R, Aune D, Overvad K, Liao L, Zhang X, Gao YT, Johansson M, Willett W, Zheng W, Shu XO. Dietary Fat Intake and Lung Cancer Risk: A Pooled Analysis. J Clin Oncol. 2017 Sep 10;35(26):3055-3064. doi: 10.1200/JCO.2017.73.3329. Epub 2017 Jul 25. PMID: 28742456; PMCID: PMC5590804.

42. Wei H, Jing H, Wei Q, Wei G, Heng Z. Associations of the risk of lung cancer with serum 25-hydroxyvitamin D level and dietary vitamin D intake: A dose-response

PRISMA meta-analysis. Medicine (Baltimore). 2018 Sep;97(37):e12282. doi: 10.1097/MD.0000000000012282. PMID: 30212966; PMCID: PMC6155959.

43. Cheng TY, Lacroix AZ, Beresford SA, Goodman GE, Thornquist MD, Zheng Y, Chlebowski RT, Ho GY, Neuhouser ML. Vitamin D intake and lung cancer risk in the Women's Health Initiative. Am J Clin Nutr. 2013 Oct;98(4):1002-11. doi: 10.3945/ajcn.112.055905. Epub 2013 Aug 21. PMID: 23966428; PMCID: PMC3778856.

44. Pai SG, Carneiro BA, Mota JM, Costa R, Leite CA, Barroso-Sousa R, Kaplan JB, Chae YK, Giles FJ. Wnt/beta-catenin pathway: modulating anticancer immune response. J Hematol Oncol. 2017 May 5;10(1):101. doi: 10.1186/s13045-017-0471-6. PMID: 28476164; PMCID: PMC5420131.

45. Takata Y, Shu XO, Yang G, Li H, Dai Q, Gao J, Cai Q, Gao YT, Zheng W. Calcium intake and lung cancer risk among female nonsmokers: a report from the Shanghai Women's Health Study. Cancer Epidemiol Biomarkers Prev. 2013 Jan;22(1):50-7. doi: 10.1158/1055-9965.EPI-12-0915-T. Epub 2012 Oct 23. PMID: 23093548; PMCID: PMC3538907.

46. Yu D, Takata Y, Smith-Warner SA, Blot W, Sawada N, White E, Freedman N, Robien K, Giovannucci E, Zhang X, Park Y, Gao YT, Chlebowski RT, Langhammer A, Yang G, Severi G, Manjer J, Khaw KT, Weiderpass E, Liao LM, Caporaso N, Krokstad S, Hveem K, Sinha R, Ziegler R, Tsugane S, Xiang YB, Johansson M, Zheng W, Shu XO. Prediagnostic Calcium Intake and Lung Cancer Survival: A Pooled Analysis of 12 Cohort Studies. Cancer Epidemiol Biomarkers Prev. 2017 Jul;26(7):1060-1070. doi:

10.1158/1055-9965.EPI-16-0863. Epub 2017 Mar 6. PMID: 28264875; PMCID: PMC5500413.

47. Riihimäki M, Hemminki A, Fallah M, Thomsen H, Sundquist K, Sundquist J, Hemminki K. Metastatic sites and survival in lung cancer. Lung Cancer. 2014 Oct;86(1):78-84. doi: 10.1016/j.lungcan.2014.07.020. Epub 2014 Aug 2. PMID: 25130083.

48. Zhou W, Park S, Liu G, Miller DP, Wang LI, Pothier L, Wain JC, Lynch TJ, Giovannucci E, Christiani DC. Dietary iron, zinc, and calcium and the risk of lung cancer. Epidemiology. 2005 Nov;16(6):772-9. doi: 10.1097/01.ede.0000181311.11585.59. PMID: 16222167.

49. Choi SH, Stommel M. Impact of Age at Smoking Initiation on Smoking-Related Morbidity and All-Cause Mortality. Am J Prev Med. 2017 Jul;53(1):33-41. doi: 10.1016/j.amepre.2016.12.009. Epub 2017 Feb 3. PMID: 28169018.

50. Bruno RS, Traber MG. Cigarette smoke alters human vitamin E requirements. J Nutr. 2005 Apr;135(4):671-4. doi: 10.1093/jn/135.4.671. PMID: 15795416.

51. Chen Q, Li L, Yi J, Huang K, Shen R, Wu R, Yao C. Waist circumference increases risk of coronary heart disease: Evidence from a Mendelian randomization study. Mol Genet Genomic Med. 2020 Apr;8(4):e1186. doi: 10.1002/mgg3.1186. Epub 2020 Feb 24. PMID: 32090477; PMCID: PMC7196469.

52. Sun H, Zheng M, Wu S, Chen M, Cai J, Yang X. Waist circumference and incidence of hypertension in Chinese adults : Observations from the Kailuan Study. Herz. 2017 Nov;42(7):677-683. English. doi: 10.1007/s00059-016-4501-x. Epub 2016 Dec 7. PMID: 27928596.

53. Vazquez G, Duval S, Jacobs DR Jr, Silventoinen K. Comparison of body mass index, waist circumference, and waist/hip ratio in predicting incident diabetes: a meta-analysis. Epidemiol Rev. 2007;29:115-28. doi: 10.1093/epirev/mxm008. Epub 2007 May 10. PMID: 17494056.

54. Hidayat K, Du X, Chen G, Shi M, Shi B. Abdominal Obesity and Lung Cancer Risk: Systematic Review and Meta-Analysis of Prospective Studies. Nutrients. 2016 Dec 15;8(12):810. doi: 10.3390/nu8120810. PMID: 27983672; PMCID: PMC5188465.

55. Halkjaer J, Tjønneland A, Thomsen BL, Overvad K, Sørensen TI. Intake of macronutrients as predictors of 5-y changes in waist circumference. Am J Clin Nutr. 2006 Oct;84(4):789-97. doi: 10.1093/ajcn/84.4.789. PMID: 17023705.

56. Calder PC. The role of marine omega-3 (n-3) fatty acids in inflammatory processes, atherosclerosis and plaque stability. Mol Nutr Food Res. 2012 Jul;56(7):1073-80. doi: 10.1002/mnfr.201100710. PMID: 22760980.

57. Jiang S, Liu Z, Wu L, Yuan Y, Hu Y, Zhang X, Wei L, Zu Y. Tumor targeting with docosahexaenoic acid-conjugated docetaxel for inhibiting lung cancer metastasis to bone. Oncol Lett. 2018 Sep;16(3):2911-2920. doi: 10.3892/ol.2018.9047. Epub 2018 Jun 28. PMID: 30127879; PMCID: PMC6096075.

58. Bai X, Shao J, Zhou S, Zhao Z, Li F, Xiang R, Zhao AZ, Pan J. Inhibition of lung cancer growth and metastasis by DHA and its metabolite, RvD1, through miR-138-5p/FOXC1 pathway. J Exp Clin Cancer Res. 2019 Nov 29;38(1):479. doi: 10.1186/s13046-019-1478-3. PMID: 31783879; PMCID: PMC6884860.

59. Liu Y, Tian Y, Cai W, Guo Y, Xue C, Wang J. DHA/EPA-Enriched Phosphatidylcholine Suppresses Tumor Growth and Metastasis via Activating Peroxisome Proliferator-Activated Receptor γ in Lewis Lung Cancer Mice. J Agric Food Chem. 2021 Jan 20;69(2):676-685. doi: 10.1021/acs.jafc.0c06890. Epub 2021 Jan 6. PMID: 33406839.

60. Bhatt SP, Kim YI, Harrington KF, Hokanson JE, Lutz SM, Cho MH, DeMeo DL, Wells JM, Make BJ, Rennard SI, Washko GR, Foreman MG, Tashkin DP, Wise RA, Dransfield MT, Bailey WC; COPDGene Investigators. Smoking duration alone provides stronger risk estimates of chronic obstructive pulmonary disease than pack-years. Thorax. 2018 May;73(5):414-421. doi: 10.1136/thoraxjnl-2017-210722. Epub 2018 Jan 11. PMID: 29326298; PMCID: PMC5903957.

61. Takamori S, Okamoto T, Mori M. ASO Author Reflections: Which of Smoking Duration or Smoking Intensity Contributes to Poor Prognosis After Resection of Non-small Cell Lung Cancer? Ann Surg Oncol. 2020 Dec;27(Suppl 3):834-835. doi: 10.1245/s10434-020-08852-5. Epub 2020 Jul 8. PMID: 32642994.