Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Ronald Eldridge Jr Date

The Novel Use of Structural Equation Modeling to Investigate
Oxidative Stress, its Determinants, and its Association with Colorectal Adenomas

By

Ronald Eldridge Jr.

Doctor of Philosophy

Epidemiology

_____

W. Dana Flanders, M.D., Sc.D.
Advisor

_____

Michael Goodman, M.D., M.P.H.
Advisor

_____

Roberd Bostick, M.D., M.P.H.
Committee Member

_____

Veronika Fedirko, PhD, M.P.H.
Committee Member

Accepted:

_____

Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

_____

Date

The Novel Use of Structural Equation Modeling to Investigate Oxidative Stress, its Determinants, and its Association with Colorectal Adenomas

By

Ronald Eldridge Jr.

B.S., Loyola University Chicago, 2008

M.P.H., Emory University, 2010

Advisor: W. Dana Flanders, M.D., Sc.D.

Advisor: Michael Goodman, M.D., M.P.H.

An abstract of
a dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Epidemiology
2014

Abstract

The Novel Use of Structural Equation Modeling to Investigate Oxidative Stress, its Determinants, and its Association with Colorectal Adenomas

By Ronald Eldridge

Despite strong basic science evidence demonstrating the role of oxidative stress in carcinogenesis, the results of epidemiologic studies addressing this issue are unconvincing. Oxidative stress is a complex, multifaceted, incompletely understood process that is unobservable *in vivo*. Numerous biomarkers of *in vivo* oxidation have been used in research, but none of these serve as a comprehensive measure. Structural equation modeling (SEM) offers the possibility of measuring oxidative stress through a latent (unobserved) variable derived from the shared covariance of multiple imperfect biomarkers, modeled in a system of *a priori* specified structural (causal) equations.

The primary objective of this dissertation was to investigate the validity and utility of the SEM method to study oxidative stress, its determinants, and its health effects.

In the first study, using three different datasets, I investigated whether a SEM would suitably identify and characterize oxidative stress from five *a priori* selected biomarkers: F_2-isoprostanes, fluorescent oxidation products, mitochondrial DNA copy number, gamma-tocopherol, and C-reactive protein. From the resulting characterization of the latent variable, and its associations with pro- and antioxidant exposures, I determined that the latent variable could be justifiably called "oxidative stress".

In the second study, I investigated the association between the latent oxidative stress variable and newly diagnosed colorectal adenoma. Based on the data from two colonoscopy-based cross-sectional studies, oxidative stress was strongly associated with colorectal adenoma with an odds ratio of 2.61 and a 95% confidence interval of 1.25-5.46 per standard deviation change in oxidative stress.

In the third study, I critically evaluated the causal assumptions of SEM when applied to studies of biologic pathways and constructs. Through multiple Monte Carlo simulations, I examined measurement error, selection bias, and unmeasured confounding using the previously examined models as case studies.

Compared to previous studies that relied on more traditional analytic techniques, the SEM method allows for a better measure of oxidative stress, and yielded a stronger association with colorectal adenoma. The methodology can be applied to other oxidative stress-related health outcomes, or possibly extended to other areas of research where it is necessary to combine different, but imperfect measurements to describe a complex biologic phenomenon.

The Novel Use of Structural Equation Modeling to Investigate Oxidative Stress, its Determinants, and its Association with Colorectal Adenomas

By

Ronald Eldridge Jr.

B.S., Loyola University Chicago, 2008

M.P.H., Emory University, 2010

Advisor: W. Dana Flanders, M.D., Sc.D.

Advisor: Michael Goodman, M.D., M.P.H.

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Epidemiology
2014

Table of Contents

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1. BACKGROUND AND RESEARCH PLAN

SECTION 1. BACKGROUND

1.1. Oxidative Stress

Reduction and oxidation (redox) are basic chemical reactions of transferring electrons from one molecule to another thereby changing their redox state. Oxidation of molecules is a normal biologic process that occurs constantly throughout cellular life. An example is cellular respiration, where glucose ($C_6H_{12}O_6$) is oxidized into water (H_2O) and carbon dioxide (CO_2) for the purpose of obtaining cellular energy. During this process, the formation of highly reactive molecules called reactive oxygen species (ROS) or reactive nitrogen species (RNS), can occur during the mitochondrial electron transport chain cascade (1). These reactive compounds have necessary roles in cellular signaling but can be damaging when not tightly controlled. For instance, hydrogen peroxide (H_2O_2) is necessary for endogenous production of thyroid hormones and acts as an intracellular messenger (2). It serves as a substrate in the iron or copper-dependent Fenton reaction ($H_2O_2 + Fe^{2+}$ (or Cu^{+}) $\rightarrow$ OH• + OH^{-} + Fe^{3+} (or Cu^{2+})), which creates hydroxyl radicals (OH•), the most reactive and damaging ROS. When ROS and RNS are "free" they can interact with cellular DNA, proteins, and other molecules resulting in mutations and/or abnormal cellular function, with one possible outcome being carcinogenesis (3, 4). Exposure to known environmental carcinogens (e.g., cigarette smoke or radiation) also leads to the formation of ROS. Protection against ROS-induced damage can be achieved through detoxification enzymes or extrinsic antioxidant molecules that reduce the ROS and return the cell to redox homeostasis (1). Many of the known protective antioxidant molecules (e.g., selenium, vitamin E, or vitamin C) are obtained through healthy human nutrition and supplementation.

Oxidative stress has long been defined as an imbalance between pro-oxidants (those that oxidize molecules and form ROS) and antioxidants (those that reduce molecules and inhibit ROS

formation) in the favor of the former (5). More recently this definition of oxidative stress was amended to account for an alternative mechanism – a disruption of thiol-redox circuits, which leads to aberrant cell signaling and dysfunctional redox control without involving ROS-induced macromolecular damage (6). This newer definition allows research to focus on specific cell types, molecular pathways and interventions but makes it difficult to examine and quantify oxidative stress in large epidemiologic studies. The larger studies tend to focus on markers of oxidation and oxidation products (disease-oriented research) or antioxidants and antioxidant enzymes (nutrition research)(6).

Evidence from basic science indicates that oxidative stress may contribute to the development of cardiovascular disease and cancer, and may influence the rate of aging (3, 7, 8). These findings provided the biologic rationale for observational studies and clinical trials of antioxidants, which yielded inconsistent results. The most notable cancer prevention trials of antioxidants were the Alpha-Tocopherol Beta-Carotene (ATBC) study, the Beta-Carotene and Retinol Efficacy Trial (CARET), the Selenium and Vitamin E Cancer Prevention Trial (SELECT), and the Women's Health Study (WHS), which tested a number of agents including β-carotene, vitamin E, vitamin C, selenium, retinol, zinc, riboflavin, and molybdenum (9). All four of these large clinical trials have resulted in null or harmful effects of the interventions on the primary outcomes of interest. There are many theories as to why the RCTs found null results; selection of high risk subjects versus population average risk subjects, incorrect dosage of intervention, poor adherence to treatment, and incorrect timing or length of treatment are some examples. However, one theory is that the complex nature of the oxidative stress causal pathway makes it difficult for the common parallel-arm or even factorial designs of RCTs to adequately represent the effects of oxidative stress-related exposures (9). The multi-factor complexities, and the biologic interactions involving these exposures, make it difficult to isolate a single causal effect (10, 11) when so many are highly correlated with each other. One proposed approach for

dealing with multiple pro- and antioxidant exposures is to create a combination variable, called oxidative balance score (OBS) (12).

The OBS is a single composite variable that is comprised of positively weighted antioxidant exposures and negatively weighted pro-oxidant exposures. The individual variables are z-score standardized or categorized based on study population distributions of each variable, weighted, then summed together to form one composite score. The weights are chosen *a priori* using a variety of approaches. Past use of OBS has resulted in larger and more consistent measures of association compared to the corresponding estimates for individual pro- and antioxidant variables with regards to all cause mortality, cancer mortality, incident colorectal adenoma, and colorectal cancer (12-17). An advantage of the OBS method is that it avoids the problem of collinearity between variables which can be problematic when modeling exposures individually. Although OBS is a useful tool for investigating multi-factor oxidative stress exposures, the method does have drawbacks. First, the variable categories and standardizations can be highly data driven, although biologic cut points can be used if known *a priori*. Second, weights for the variables are either completely arbitrary, derived from previous literature which tends to be conflicting, or derived from regression coefficients which are completely data driven. None of these three methods provides the researcher with pre- or post-analysis evidence of validity. Third, any interactions involving score components are usually not taken into account because one must incorporate interactions into the score variable, and also specify whether those interactions are synergistic or antagonistic; and this information is rarely, if ever, available. And lastly, because the OBS includes many behavioral characteristics such as smoking and alcohol drinking history, medication use, and sometimes even physical activity, there is little to no guarantee that such a composite variable is representing oxidative stress wholly or even partially, rather than just healthy lifestyle (18). Even if the OBS is representing oxidative stress, it most

certainly is also representing other biologic mechanisms involved in carcinogenesis (19-25). Separation of these mechanisms is impossible using the OBS method.

Another method of studying the effects of multiple factors that may act through a common pathway or pathways is to use structural equation modeling with a latent (unmeasured) variable. The following sections of this communication discuss this methodology in general and specifically how it can be applied to a study of oxidative stress and colorectal adenoma.

1.2. Structural Equation Modeling

Structural equation modeling (SEM) is a multivariate modeling technique combining path and measurement analyses. This methodology requires that researchers specify the structural (causal) relationships between modeled variables, both observed and latent, and then allows estimating those causal parameters from non-experimental data under certain assumptions. Although its origins date back to the early 1900s with published works by Sewell Wright on path analysis and Charles Spearman on factor analysis, formal application of the technique was not seen until the 1980s with publication of textbooks and computer programs(26-28). The power of SEM resides in its ability to estimate multiple parameters at once, while simultaneously reducing measurement error through the creation of latent variables.

Path analysis is a general term used to describe any multivariate regression technique applied in the assessment of causal/structural dependencies between manifest variables. A path model is commonly but not exclusively depicted as a causal directed acyclic graph (cDAG) where the arrows represent the causal effects and the nodes represent the variables. A simple Markovian example is shown below (Figure 1.1), where A causes B, and where both A and B cause C and where the U_i are errors/disturbances. The variable A is called an exogenous variable because its causes (other than error) are not depicted in the model, while B and C are called endogenous variables, because their causes are depicted in the model (26, 27). Each arrow is a direct effect

from one variable to another and the path A-B-C is an indirect effect of A on C through the mediating variable B. The total effect of A on C would be the direct effect added to the indirect effect. Path analysis simultaneously estimates the effect of each of the arrows along with overall model fit (26, 27). Identifiability is a term used to describe the ability to estimate unbiased direct and indirect effects. Requirements for identifiability are no unmeasured confounding of the total effect and of the intermediate and outcome effect (29-31).

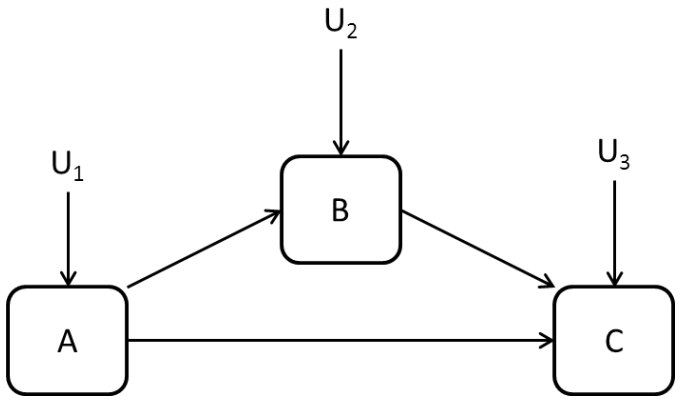

Figure 1.1. A simplified path model of three variables

The term "measurement analysis" describes any multivariate technique used to combine multiple correlated variables into a single latent variable representing an unmeasured construct. Common measurement analysis methods are confirmatory or exploratory factor analysis (CFA, EFA) and principle components analysis. The difference between CFA and EFA are *a priori* decisions as to how many latent factors to retain, and the relationship between the predictors and the factors (26, 27). The measurement model is depicted by observed/manifest variables, usually represented by squares or rectangles that are caused by unobserved/latent variables, usually represented by circles or ovals. A simple example is shown below (Figure 1.2) where A is the latent variable and is predicting three manifest variables, B, C and D with corresponding error

terms. The diagram below is an example of assuming local independence - where the correlation between B, C, and D is exclusively explained by the latent variable A. The arrows between the latent and manifest variables are called factor loadings.

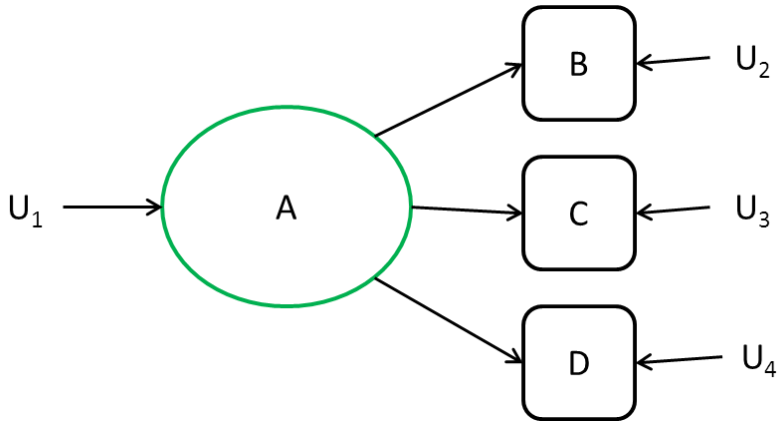

Figure 1.2. A simplified measurement model of three observed and one latent variable

SEM is a powerful analysis tool, but it does have limitations and requires a number of assumptions. First, the researcher must assume causal relationships between variables although the data to support causality (or at least temporality) may not be available. Additionally, every path included and excluded is based on explicit or implicit assumptions about the direct effects, indirect effects, and confounding associations. SEM allows for easy estimation of direct and indirect effects and although the assumptions are well established (29-34), there is much less literature on when and where these assumptions are plausible and possible solutions when they aren't (32, 33). In addition, SEM has statistical weaknesses by forcing a researcher to assume an underlying distribution for every endogenous variable rather than just a single outcome as is done in normal regression methods. Additionally, the simultaneous estimation method can cause convergence problems and an inability to estimate parameters, especially in smaller data sets. Finally, as latent variable's representation of the unmeasured construct is solely dependent upon

the selection of the indicator variables, shared error of the predictors will result in residual error in the latent variable.

The main strength of SEM is its multivariate properties that allow estimating numerous effects simultaneously. SEM is also well suited for the calculation of direct and indirect effects provided that the underlying causal structure is correct (35). Although there are concerns about calculations of direct and indirect effects, sensitivity analyses may help assess the robustness of conclusions (32, 33). SEM is also well suited for inclusion of latent variables. The technique can reduce error through measurement analysis while estimating the effects on and from a latent construct. Lastly, the multivariate nature allows for relative statistical tests of nested models and absolute tests of model fit, both overall and with regards to the latent variable(26, 27).

1.3. Oxidative Stress Biomarkers

Most ROS, especially the hydroxyl radical (OH•), are difficult to measure directly. The available methods of directly measuring ROS (electron spin resonance or spin trapping) are impractical for large epidemiologic studies (36). Therefore, epidemiologists are more likely to measure the end products of ROS interactions with macromolecules. These end products of oxidation can be similar or can differ (6); they can be specific (37) or general (38); but they all are imperfect measures with regards to oxidative stress. There are numerous biomarkers; of those, F_2-isoprostanes, fluorescent oxidation products, mitochondrial DNA copy number and γ-tocopherol, are available for analyses outlined in the present proposal.

F_2-isoprostanes (FIP) are prostaglandin-like compounds that are formed by the free-radical peroxidation of arachidonic acid (Figure 1.3). FIP are considered a reliable and specific biomarker of lipid peroxidation that can be measured in either blood plasma or urine. FIP concentration is considered the gold standard *in vivo* biomarker of lipid peroxidation because it is specific, stable compared to malondialdehyde, demonstrates a dose-response relationship with

oxidant injury, and is independent of dietary lipid content (37, 39, 40). Despite the biologic evidence linking oxidative stress to colorectal neoplasia, studies evaluating the association between FIP and adenomatous polyps reported inconsistent results (41, 42). FIP have been shown to be positively associated with factors assumed to affect oxidative stress but have mixed results with regards to dietary antioxidants (40, 42-44). However, a recent finding from a pilot clinical trial showed that an antioxidant cocktail of α-tocopherol, β-carotene, vitamin C, selenomethionine, riboflavin, niacin, zinc, and manganese, could have differential effects on FIP by smoking status with decreased levels in non-smokers but increased levels in smokers (45).

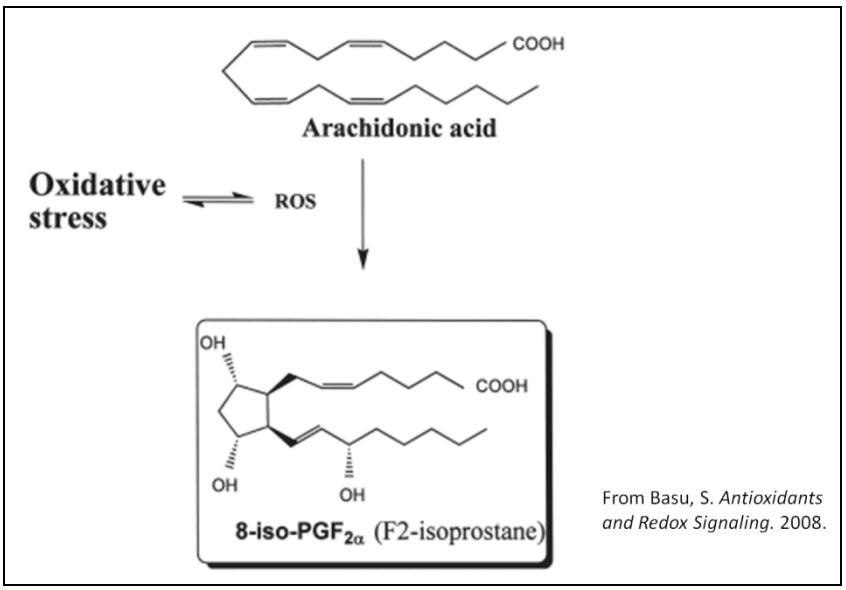

Figure 1.3. F_2-isorprotanes formation from arachidonic acid (40)

Plasma fluorescent oxidation products (FOP) measure oxidized DNA, proteins, and lipids (46, 47). This non-specific mixture of oxidation end-products only recently was introduced as a possible biomarker for oxidative stress (38). The assay is very stable (39) and correlates with factors and conditions known to be associated with differing concentrations of oxidative stress markers (creatinine, total cholesterol, smoking, hypertension, CVD) (38). The presumed

advantage of FOP compared to FIP is that it reflects oxidation of all macromolecules (including DNA), not just lipids (48). The main disadvantages of FOP as a biomarker include the relative scarcity of data supporting its use and the poor understanding of the underlying biochemical mechanisms that result in its production *in vivo* (38).

Mitochondrial DNA (MtDNA) is the double-stranded DNA in the human mitochondria that code for proteins responsible for cellular respiration and apoptosis (49, 50). Each cell has multiple mitochondria and each mitochondrion has multiple copies of DNA. Among healthy cells, the copy number of nuclear DNA is fairly stable (49). By contrast, MtDNA copy number is highly variable because MtDNA has limited repair capacity and is highly susceptible to damage due to its close proximity to the ROS created by the electron transport chain (49, 50). MtDNA has been implicated in the carcinogenic process and there is good evidence that MtDNA is altered as a cause or result of cancer, including colorectal carcinoma (49-54). One recent study found an increased risk of colorectal cancer in individuals with high pre-diagnosis MtDNA copy number (55). However there are limited studies relating MtDNA copy number to other biomarkers of oxidative stress (56) or to colorectal adenoma.

Gamma-tocopherol (Gtoc) is one of eight structurally related forms of vitamin E. Chemically, Gtoc is an antioxidant with a slightly lower potency than that of alpha-tocopherol (57). However, there are mixed results regarding the effects of Gtoc supplementation on oxidative stress biomarkers with some studies finding positive associations (58-60) and others finding null (61). More importantly, the non-supplementation plasma levels of Gtoc may in fact be a marker of oxidative stress as they are positively associated with FIP and C-reactive protein (CRP) (62). The current theory is that the primary metabolite of Gtoc, γ-carboxyethyl-hydroxychroman (γ-CEHC), may mediate the antioxidant and anti-inflammatory effects of Gtoc (60), and that the degradation of Gtoc by cytochrome P450 can be altered by oxidative stress (57). Therefore, plasma levels of Gtoc may in fact be a better representation of underlying conditions rather than

dietary intake. With regards to colorectal adenoma, Ingles et al. reported a positive association between higher plasma Gtoc levels and adenoma (63).

1.4. Relation between Oxidative Stress and Inflammation

Chronic inflammation is closely linked with oxidative stress (3, 4, 64, 65). The mechanism behind ROS formation and inflammation is not completely understood. It is recognized that inflammatory leukocytes can undergo a "respiratory burst" at the site of tissue damage or infection ROS levels resulting in localized oxidative stress (65). Conversely, consistently low to moderate levels of ROS can increase production of the nuclear transcription factor NF-κβ, which in turn increases the production of inflammatory cytokines (3, 65). However, higher levels of ROS tend to inhibit NF-κβ (65). Despite the lack of complete knowledge of how ROS and inflammation affect each other, it is clear that they are closely related.

C-reactive protein (CRP) is an acute-phase, non-specific blood protein that rises in response to pro-inflammatory stimuli. A cross-sectional study has shown that lifestyle factors correlated with oxidative stress are also similarly correlated with CRP (43). Moreover, CRP has been positively correlated with oxidative stress biomarkers: circulating oxidized low density lipoprotein, oxidized mononuclear cells, free-oxygen radical test (FORT) results, and urinary FIP (66-69). Other oxidative stress biomarkers, such as the glutathione/glutathione disulfide ratio (GSH/GSSG), are not correlated with CRP, an observation suggesting that different biomarkers represent different oxidative processes with variable relation to inflammation (68). Thus CRP may be linked to lipid oxidation (as measured by FIP & FORT) but not to thiol oxidation process (as measured by GSH/GSSG) (68). Serum level of CRP is frequently used as a predictor of cardiovascular disease as well as atherosclerosis (70, 71). There is also evidence to suggest it could be a marker of inflammation in the colon, particularly for Crohn's disease, and to a lesser extent for ulcerative colitis (72). In a meta-analysis, CRP is reported to be weakly associated with

colorectal cancer incidence (73-75). In contrast, prospective studies do not suggest an association between CRP and colorectal adenoma (76-78).

SECTION 2. DISSERTATION RESEARCH PLAN

2.1. Objectives

The primary objective of this proposal is to explore the novel use of a SEM with a latent variable to investigate a simplified causal pathway of oxidative stress to colorectal adenoma (Figure 1.4). Using pooled cross-sectional studies, I will model a simplified pathway of pro- and antioxidant exposures, oxidative stress and colorectal adenoma. Oxidative stress will be a continuous latent variable from the combined correlation between four biomarkers measuring products of oxidation (FIP, FOP, MtDNA, and Gtoc) and one biomarker measuring inflammation (CRP). This latent outcome variable will also be used as the exposure of interest for colorectal adenoma risk. Adenoma status will be a dichotomous variable determined through routine colonoscopy. The validity of the latent variable will be assessed along with the assumptions and potential biases. The overall objectives of this dissertation will be achieved by addressing the following specific aims.

2.2. Specific Aims

Aim 1: To assess whether a latent variable constructed from FIP, FOP, MtDNA, Gtoc, and CRP will suitably identify and characterize oxidative stress.

Aim 2: To investigate whether the latent oxidative stress variable is associated with colorect adenoma.

Aim 3: To critically evaluate the causal assumptions and quantitatively investigate the potential biases of the SEM method as it applies to studies of biologic phenomena.

Hypotheses

We hypothesize that most individual OBS components will have moderate effects on the latent oxidative stress variable. We also hypothesize that oxidative stress latent variable will have a small, to moderate effect on colorectal adenoma. We don't expect the assumptions and biases to drastically alter our conclusions from either the first or second aim.

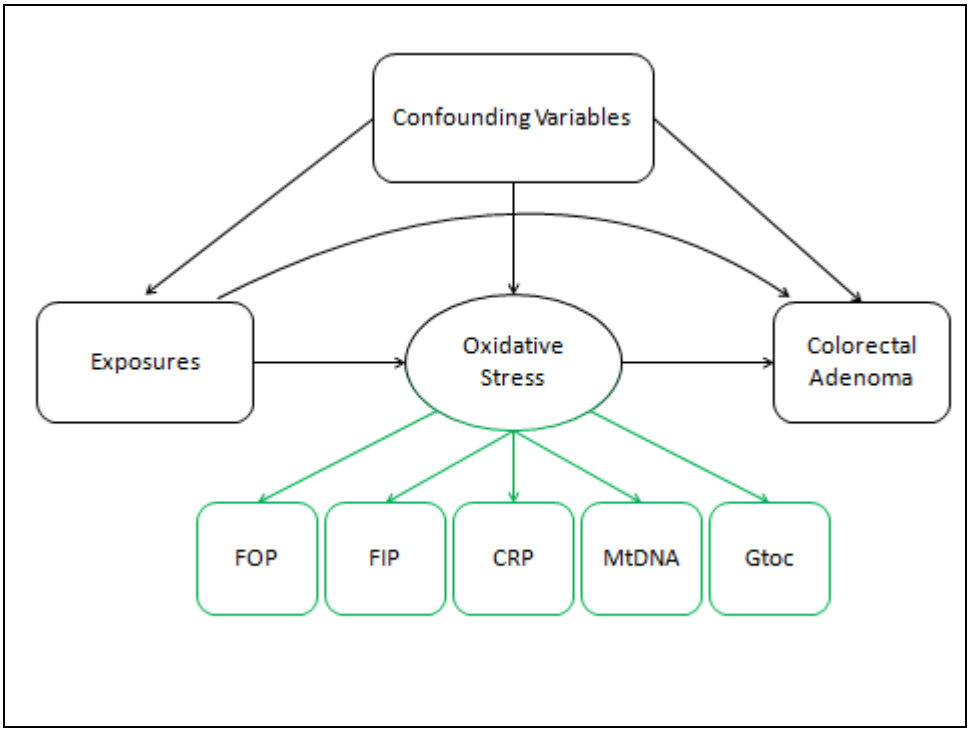

Figure 1.4. The theorized model of the latent oxidative stress variable and colorectal adenoma

2.3. Data Sources

Aim 1

Aim 1 will use three pooled study populations: non-cases from the Markers of Adenomatous Polyps (MAP) studies I and II (n=707) (described in aim 2), and participants in the

cross-sectional Study of Race, Stress, and Hypertension (SRSH). SRSH is a pilot study that recruited approximately 324 subjects from three racial /ethnic groups: Whites (n=124), African Americans (n=99), and West African immigrants (n=101). The White and African American subjects were a random sample from the Georgia Cancer Study (GCS), a pilot state-wide cohort study of approximately 800 participants recruited in 2007-08. West African immigrants were recruited *de novo* in 2011 from Atlanta churches. The recruitment and data collection protocol for the West African subjects followed that of the GCS. Consented subjects filled out questionnaires about medical history and demographics but did not provide dietary or supplement information. All subjects also underwent anthropometric and blood pressure measurements. Blood was drawn, processed, and analyzed for all dietary and oxidative stress biomarkers in the same laboratories using the same methods as was done in the MAP I and II (described in aim 2).

Aim 2

The data used to investigate aim 2 will be pooled from two similarly conducted cross-sectional studies of newly diagnosed colorectal adenoma: the Markers of Adenomatous Polyps I and II (MAP I and MAP II), for a combined study population of 707 persons of whom 233 had were diagnosed with colorectal adenoma and 312 were controls (162 missing adenoma status) (15). The subjects for the two MAP studies were recruited from community gastroenterology practices in Winston-Salem and Charlotte, North Carolina and in Columbia, South Carolina. Patients with no prior history of colorectal neoplasms were invited to participate in the study when they were scheduled to undergo elective colonoscopy. Before their colonoscopy, all consenting subjects completed mailed questionnaires that collected information on their demographic characteristics, medical history, and habits. Diet and use of supplements were assessed through a modified Willet food frequency questionnaire (FFQ). In addition, all participants provided blood samples that were processed, stored at -70°C, and analyzed for dietary and oxidative stress biomarkers by high-performance liquid chromatography with the

exception of MtDNA copies. The MtDNA biomarker was analyzed in 2013 by Dr. Thyagarajan's laboratory at the University of Minnesota using real time quantitative PCR. The complete laboratory method has been previously published (55). Cases were defined as subjects diagnosed as having at least one colon or rectal adenomatous polyp. Controls were subjects free from all adenomas. Details of the data collection and laboratory methods have been previously published (79-81). Nearly all MAP study participants were non-Hispanic whites.

Aim 3

The critical evaluation of SEM assumptions will be done theoretically through directed acyclic graph theory and will therefore not require data sources (82). The quantification of potential bias will be done by simulated populations. Using the Monte Carlo simulation feature in the Mplus statistical program (83), I will create 500 populations of 1,000 subjects each for every combination of specified parameters. The ausal parameters of interest will be set while biasing parameters will be altered to provide a potential range of bias. I will consider potential bias from information, selection, and confounding.

2.4. Research Plan

Aim 1

The following section explains the research plan for assessing the suitability of the latent oxidative stress variable. It first describes the latent variable, and then the exposures used to validate the latent variable. It also describes the tests of invariance that will be conducted.

Oxidative stress Latent Variable

The oxidative stress variable in this analysis is composed of one latent factor indicated by five continuous blood biomarkers: FIP, FOP, MtDNA, CRP, Gtoc (Figure 1.5). The indicators will be tested for normality and log transformed if needed. The latent factor will be scaled by

standardization. Assumptions are that the residuals of the indicators are Gaussian, have an expected value of zero, are uncorrelated with each other (local independence) and are uncorrelated with the latent factor. The following are the matrix equations and the diagram of the latent factor (η):

$$y_i = \nu_i + \Lambda_i \eta + \varepsilon_i \quad \text{where } y_i = \begin{bmatrix} FIP \\ FOP \\ CRP \\ MtDNA \\ Gtoc \end{bmatrix} \nu_i = \begin{bmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \\ \nu_4 \\ \nu_5 \end{bmatrix} \Lambda_i = \begin{bmatrix} \lambda_{FIP} \\ \lambda_{FOP} \\ \lambda_{CRP} \\ \lambda_{MtDNA} \\ \lambda_{Gtoc} \end{bmatrix} \varepsilon_i = \begin{bmatrix} \varepsilon_{FIP} \\ \varepsilon_{FOP} \\ \varepsilon_{CRP} \\ \varepsilon_{MtDNA} \\ \varepsilon_{Gtoc} \end{bmatrix}$$

$$E(\eta) = \alpha = 0; \quad Var(\eta) = \Phi = 1;$$

$$Cov(\varepsilon_i) = \Theta = DIAG(\theta_{FIP}, \theta_{FOP}, \theta_{CRP}, \theta_{MtDNA}, \theta_{Gtoc});$$

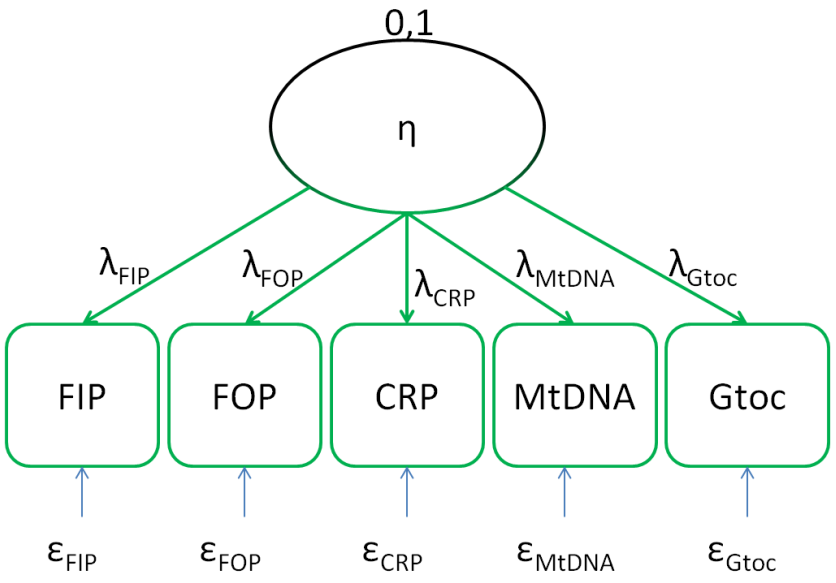

Figure 1.5. A Depiction of the oxidative stress latent variable

Exposures

Two sets of exposures will be used to investigate their associations with the latent variable. The primary set of exposures will consist of age, sex, race, body mass index (BMI),

smoking history, drinking history, regular aspirin use, and regular non-steroidal anti-inflammatory (NSAID) use. Of these, age, BMI, smoking, and drinking are theorized to increase oxidative stress, while aspirin and NSAID use are theorized to decrease oxidative stress. The secondary set of exposures will consist of plasma measures of α-carotene, β-carotene, α-tocopherol, β-cryptoxanthin, lycopene, and lutein. All are theorized to decrease oxidative stress. The primary exposures will be modeled without the secondary exposures due to the possibility that some plasma antioxidants are intermediates between a primary exposure and oxidative stress.

Invariance testing

Before pooling study populations, I will assess invariance of the latent variable. Multiple groups SEM (MG-SEM) analyses, a form of interaction testing, will be performed using non-cases subjects from the MAP study and the SRSH data. MG-SEM is a simultaneous estimation of the model among *G* number of groups, similar to that of stratified Cox models (26, 27). The MG-SEM will be used to test invariance with regard to the four biomarker factor loadings (FIP, FOP, MtDNA, CRP, Gtoc) on the single latent factor model, and then assess the regression parameters of the exposure variables. Levels of invariance fall under four categories of increasing strength: configural, weak factorial, strong factorial, and strict factorial (26, 27). Each level denotes added equality restriction(s) placed on the model to test where there might be statistical differences in the latent factor with regards to the *G* groups.

Configural invariance places no restrictions and just demonstrates that the general model can be replicated. Weak factorial invariance establishes equal factor loadings by testing the equality $\Lambda^{(0)}_{xxx} = \Lambda^{(1)}_{xxx}$ while letting all other parameters vary. Strong factorial invariance establishes equal factor loadings and intercepts by testing the equalities $\Lambda^{(0)}_{xxx} = \Lambda^{(1)}_{xxx}$ and $\nu^{(0)}_{xxx} = \nu^{(1)}_{xxx}$ while letting the indicator residuals vary. Strict factorial invariance is met by testing the loadings, intercepts, and residual variances, $\Lambda^{(0)}_{xxx} = \Lambda^{(1)}_{xxx}$; $\nu^{(0)}_{xxx} = \nu^{(1)}_{xxx}$; $\theta^{(0)}_{xxx} = \theta^{(1)}_{xxx}$ (26,

27); The "build up" strategy will be employed by starting with configural invariance and increasing the number of equalities until significant differences are found. All tests of significance will be determined by χ^2 tests along with consideration to relative improvement among generally accepted SEM fit statistics (CFI, TLI, RMSEA, SMR)(26, 27). Once the level of invariance is determined between the two datasets, differences in regression parameters can be tested and interpreted. The following diagrams are depictions of invariance testing of the oxidative stress latent variable between two groups.

Variables that will be tested for SEM group differences will draw comparisons by study data source, sex, and race in an attempt to maximize stratified group populations. The following tests will be performed:

1) Non-Hispanic Whites versus African Americans versus West Africans in SRSH
2) SRSH versus MAP datasets
3) Men versus Women

The statistical analyses for aim 1 will be performed using the latest version of Mplus software. The analyses will run on the dataset(s) specified and use the default setting of missing at random (MAR) for all missing data on any of the variables. All tests of statistical significance will be evaluated at the <0.05 p value level.

Aim 2

The following section explains the research plan for investigating the association between the latent oxidative stress variable and colorectal adenoma. It describes the structural model for estimating the effects, and the sub-analyses. The statistical plan for the latent variable is the same as in aim 1 and is presented there.

Structural Model

The proposed structural model for aim 2 includes two endogenous variables (colorectal adenoma and oxidative stress) and a yet undetermined number of exogenous variables made up of exposures and potential confounders. Variables that will be considered are the primary and secondary exposures in aim 1. In addition, variables not available or not considered in the SRSH dataset such as, vitamin C, fiber, total energy intake, poly-unsaturated fatty acid intake, physical activity, plasma cholesterol, and family history of colorectal cancer will be considered. The theoretical model of Figure 1.4 depicts the potential confounding variables as causes of the exposure variables, however these relationships will be modeled as binary covariance for statistical purposes (shown in Figure 1.6). This change removes the need to specify an underlying statistical distribution for each of the exposure variables and will have no ramifications on the estimation of the causal effects on oxidative stress or colorectal adenoma. The following are the equations and diagram of the modified structural model:

$$\eta = \alpha_1 + \gamma_{Ei1}X_{Ei1} + \gamma_{Ci1}X_{Ci1} + \varepsilon_{\eta}$$

$$logit\,[\mathrm{CA}] = \alpha_2 + \beta_{21}\eta + \gamma_{E2i}X_{Ei2} + \gamma_{Ci2}X_{Ci2} + \varepsilon_{CA}$$

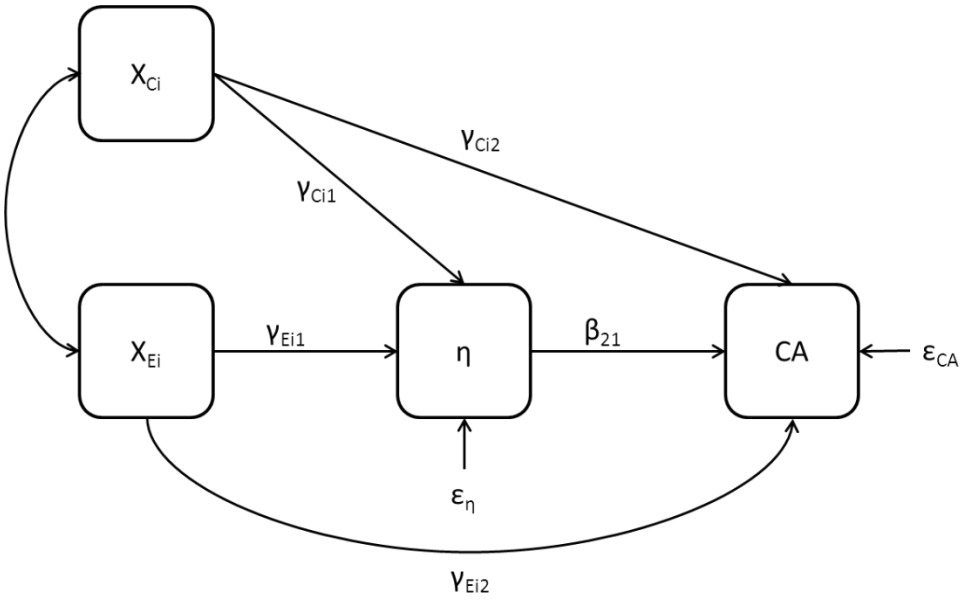

Figure 1.6. A depiction of the structural model used in aim 2.

The above model is identified by the two-step rule (26, 27): first for the latent variable, then for the structural portion. When structural portion of the model is identified any identified latent variables are then assumed to be manifest variables. The latent variable portion of my model is identified by the three-indicator rule (26, 27) – that any single latent variable can be identified with at least 3 indicators. The structural portion of the model is identified by the recursive rule (26, 27) – that the model is acyclic and has no correlated error terms (i.e., Markovian).

Sub-analyses

An alternative theory is posited to include two latent variables (η_1 and η_2) representing two distinct phenomena of oxidation. The two variables will be indicated by the same four biomarkers and will be allowed to co-vary (double-headed arrow) because there is no a priori reason to believe one latent factor would cause the other. The structural model will include the

addition of causal arrows from all exposure and confounding variables to the second latent variable in addition to a causal arrow from the second latent variable to colorectal adenoma. The two theories will be qualitatively evaluated for plausibility by assessing meaningful predictors of each latent factor. Ideally, each predictor should not have large and meaningful factor loadings on more than one latent variable. A quantitative evaluation between the two models will be performed by a χ^2 test with multiple degrees of freedom.

The alternative model's identification rules are not as simple as the main model. Because there are four indicators for two latent variables, which can meet the requirements for the two-indicator rule (26, 27), and because the model is still Markovian, we expect but are not certain that the model will be identified. Even if the model is identified, the estimation may not converge to a solution. If either does not happen, then the alternative model will be excluded from the analyses.

The statistical analyses for aim 2 will be performed using the latest version of Mplus software (83). The analyses will run on the full study dataset and use the default setting of missing at random (MAR) for all missing data on any of the variables. All tests of statistical significance will be evaluated at the <0.05 p value level.

Aim 3

The following section explains the research plan for assessing the causality assumptions and potential biases of SEMs used in studies of biologic phenomena. Throughout this aim, I will use the studies of aim 1 and aim 2 as real world example case-studies.

Causality Assumptions

The causality assumptions will be qualitatively assessed through DAG theory (82). The focus will be on the assumptions centered around the latent variable, and the potential violations

of those assumptions. I will consider the potential ramifications due to violations of an exposure variable directly causing a biomarker (not mediated by oxidative stress), a biomarker sharing a common cause with adneoma (not oxidative stress), and two biomarkers sharing a common cause (not oxidative stress). These three potential violated assumptions are shown in Figure 1.7 with dotted lines.

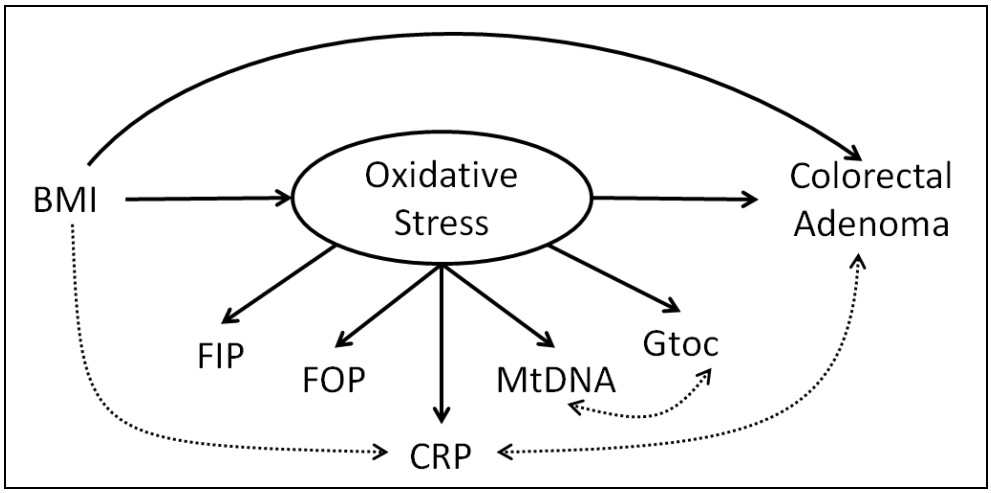

Figure 1.7. SEM assumptions and potential violations.

Potential Biases

Potential bias by sources of information, selection, and unmeasured confounding will be assessed by Monte Carlo data simulation. For each type of bias, there will be a simulated model and an analysis model. The simulated model will simulate the potential bias in the target population, while the analysis model will be model the theoretical study (biased model). To simplify the model, in each scenario, oxidative stress will be treated as a measured variable, rather than a latent one.

The investigation of information bias will focus on the temporality assumption made in aims 1 and 2, and will consider non-differential and differential bias of the effect of oxidative stress on adenoma, and the effect of BMI on oxidative stress. The non-differential bias will

assume the direction of causality as theorized in aim 2 (Figure 1.8), while the differential bias will consider the possibility of adenoma affecting oxidative stress (Figure 1.9). To model these biases, measured oxidative stress must be related to a causally relevant oxidative stress measurement; the same applies for BMI. This will be done by either direct causes in the case of BMI, or by a common cause in the case of oxidative stress. The correlation, and thereby the resulting bias, will be changed by altering the strength of the effects. The differential bias model will also include a direct effect of adenoma to measured oxidative stress. Values of the parameters will be chosen and data simulated to provide a potential range of bias. The analysis model will model only variables at time point 3 (Figure 1.8).

The investigation of selection bias will focus on disease-free survival to study inclusion. This will be modeled as conditioning on disease-free survival that is caused by both the exposure and an unmeasured risk factor for colorectal adenoma (Figure 1.10). The causal effects of interest are the effects of exposure to oxidative stress, exposure to adenoma, and oxidative stress to adenoma; these will be set. The effects and conditions that will be changed to give a range of bias will be, the effect of exposure on survival, unmeasured risk factor on survival (inverse of exposure), the unmeasured risk factor on adenoma, and the survival probability. The analysis model will condition on survival (S=1) and will model exposure, oxidative stress, and adenoma.

The investigation of unmeasured confounding will focus on unmeasured confounding of the effect from oxidative stress to colorectal adenoma. This will be modeled as an unmeasured common cause of both variables (Figure 1.11). The effects of exposure on oxidative stress, exposure on adenoma, and oxidative stress on adenoma will be set. The effect of the unmeasured variable on oxidative stress and on adenoma will be changed to provide a potential range of bias. The analysis model will only model exposure, oxidative stress, and adenoma.

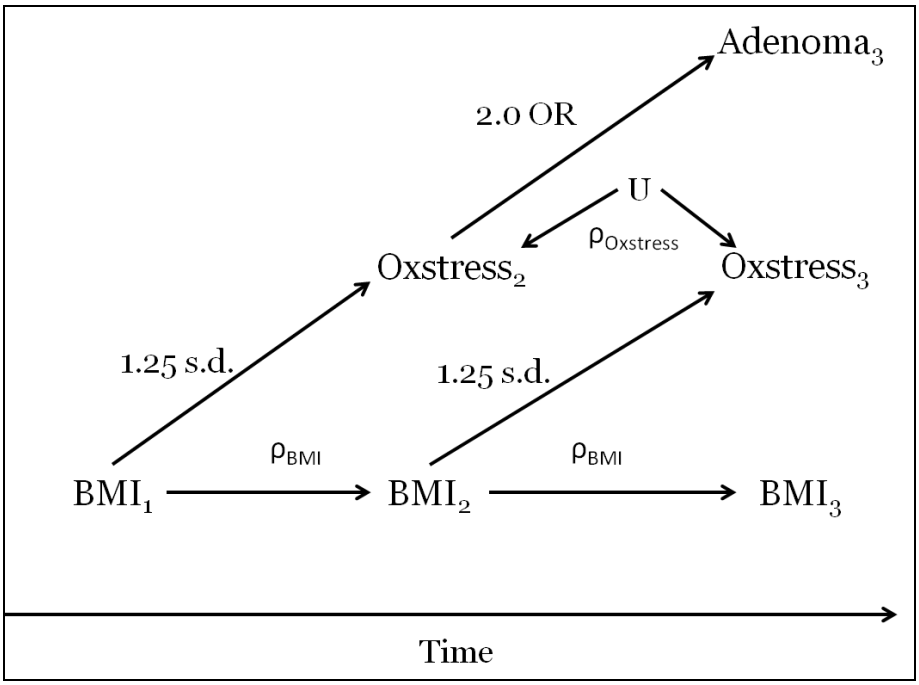

Figure 1.8. Non-differential information bias of BMI, oxidative stress, and colorectal adenoma

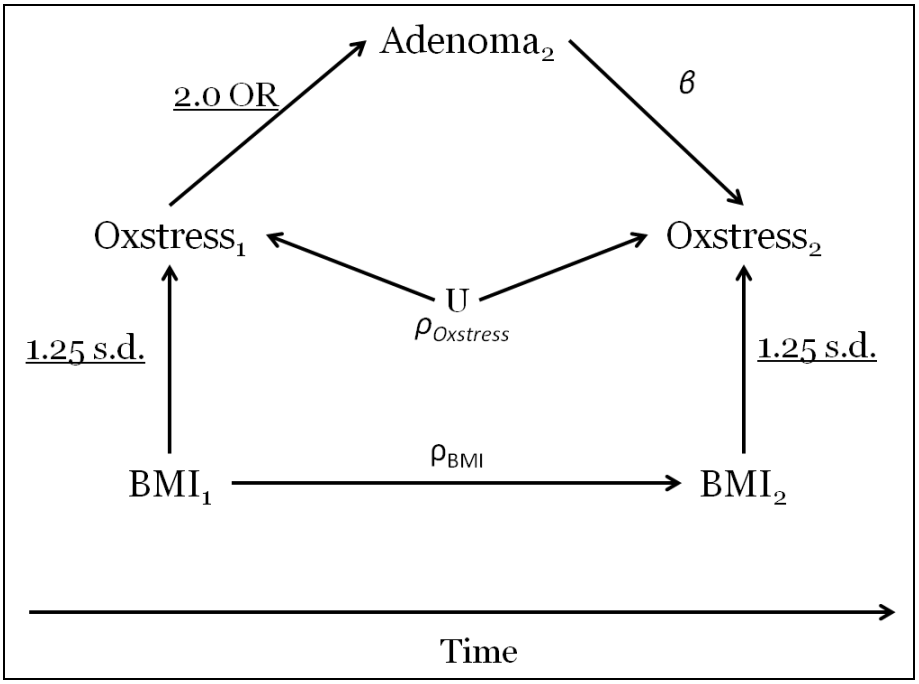

Figure 1.9. Differential information bias of BMI, oxidative stress, and colorectal adenoma

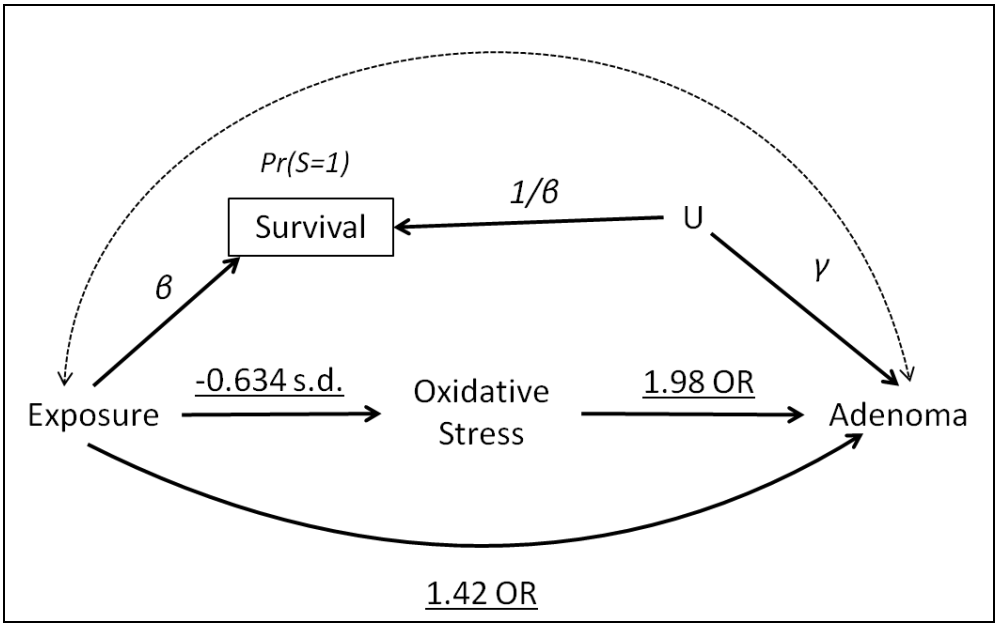

Figure 1.10. Selection bias due to disease-free survival to study inclusion

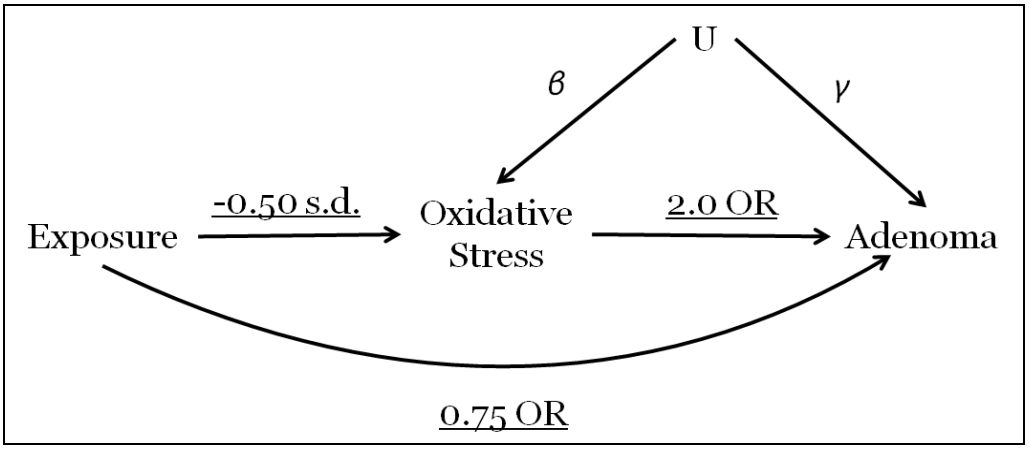

Figure 1.11. Unmeasured confounding of oxidative stress on adenoma effect

SECTION 3. NOVELTY, SIGNIFICANCE, AND IMPACT

Colorectal cancer is the 2nd most common invasive malignancy in the United States (84). Early detection and removal of pre-cancerous lesions is a major reason for decreasing colorectal cancer incidence and mortality in the past 20 years (84). Nearly 95% of all sporadic cases of colorectal cancer develop from adenomatous polyps and the lifetime risk for developing a polyp is approximately 20% (85). Many Risk factors for colorectal cancer are thought to be modifiable with changes in behavior and diet (low physical activity, obesity, high meat consumption, low fruit and vegetable consumption) and this is an important area of research (86).

There were seven studies that investigated the association between oxidative stress biomarkers and colorectal neoplasia (41, 45, 73, 76-78, 87). In addition a number of studies investigated the effects of antioxidants on oxidative stress or colorectal cancer/adenoma risk (9, 16, 40, 44, 88-92). Despite the numerous studies, the novelty of the proposed dissertation is the application of the SEM with the oxidative stress latent variable. To date, no published studies of oxidative stress biomarkers used as indicators for a latent variable are available and there are very few published studies of applying a SEM to model a simplified cancer pathway.

The benefits of applying SEM to a simplified cancer pathway include simultaneous estimation of a latent biomarker variable and testing of the underlying causal theory. If for instance, the oxidative biomarkers are formed by a latent construct that is not an intermediate between the oxidative exposure variables and colorectal adenoma, one of two things will likely happen: 1) the latent construct will be formed with high correlations with the indicators and low regressions with the exposure variables and colorectal adenoma, or 2) the latent construct will be formed with low correlations with the indicators and higher regressions with the exposure variables. By including the additional information of the theorized pathway, the study data will provide statistical evidence as to whether the latent biomarker is in fact an intermediate.

Additionally, if the latent biomarker is an intermediate, the SEM can add useful new information to the existing data on the relationship of the biomarkers to each other (38-41, 68). The general method of SEM can be applied to other areas of cancer epidemiology as long as there are viable theoretical models of the causal pathway and the ability to measure relevant biomarkers.

The proposed studies may generate new information about specific pro- and antioxidants and their effect on oxidative stress and provide clarification about the presence or absence of a causal link between oxidative stress and colorectal adenoma. The use of SRSH, in addition to MAP, will help assess if the effects of oxidative stress-related exposures on oxidative stress differ by sex or race/ethnicity. The evaluation of causal assumptions and potential biases proposed in aim 3 may improve the methodological understanding of the SEM method and its applicability to studies of biologic phenomena.

CHAPTER 2. USING MULTIPLE DETERMINANTS AND BIOMARKERS TO OBTAIN A BETTER MEASUREMENT OF OXIDATIVE STRESS: A LATENT VARIABLE STRUCTURAL EQUATION MODEL APPROACH

ABSTRACT

Background: Oxidative stress is a complex, multifaceted biologic process measured by imperfect biomarkers. No single biomarker can completely represent oxidative stress because each biomarker reflects only some aspects of this phenomenon. The analytical technique of structural equation modeling (SEM) can provide a solution to this problem by modeling a latent (unobserved) oxidative stress variable, constructed from the covariance of multiple biomarkers.

Methods: Using three pooled cross-sectional datasets, we modeled a latent oxidative stress variable from five biomarkers: F_2-isoprostanes (FIP), fluorescent oxidation products (FOP), mitochondrial DNA copy number (MtDNA copy number), γ-tocopherol (Gtoc), and C-reactive protein (CRP). We validated the latent oxidative stress variable by assessing its relation to pro-/antioxidant exposures, and illustrated the utility of this method.

Results: FIP, Gtoc, and CRP primarily characterized the latent oxidative stress variable. Obesity, smoking, aspirin use, and β-carotene were statistically significantly associated with oxidative stress in the theorized directions. The same exposures were weakly and inconsistently associated with the individual biomarkers.

Conclusions: Our results suggest that the use of SEM with latent variables decreases the individual biomarker-specific variability and may produce a better measure of oxidative stress. This methodology can be applied to similar situations, in other areas of research, where it is necessary to combine different, but imperfect measurements to describe a complex biologic phenomenon.

INTRODUCTION

The biologic process of oxidative stress is complex and incompletely understood. It is commonly defined as a disruption in the pro- and antioxidant balance, in favor of the former (5). Basic biology studies indicate that the pro-/antioxidant imbalance leads to increased production of reactive oxygen species (ROS), and that higher levels of ROS may in turn contribute to the initiation, promotion, and progression of carcinogenesis (2-4, 65, 93, 94). Contrary to the existing *in vitro* evidence, findings from human observational studies are inconsistent (90, 95-99), and randomized clinical trials of supplementary antioxidants have reported null or even harmful effects (9).

Certain lifestyle choices and demographic characteristics are thought to influence oxidative stress. Tobacco smoke is a known pro-oxidant (100). Pure alcohol is recognized as a pro-oxidant (101, 102), but as alcoholic beverages contain polyphenols and other antioxidant compounds (103, 104), and the net effect of alcohol consumption is unclear (43). Increased age and higher levels of obesity are also associated with higher levels of oxidative stress (105-107). On the other hand, aspirin use, vitamin C, vitamin E, and carotenoids are thought to act as antioxidants, either directly (by scavenging ROS), or indirectly (by assisting endogenous antioxidant enzymes), or both (1, 3, 88, 108, 109). Despite any inconsistent findings, these factors are thought to demonstrably influence *in vivo* oxidation and are used to validate biomarkers of oxidative stress (38, 40, 43, 110).

Because oxidative stress cannot be directly observed in humans, many biomarkers of this process have been proposed. The epidemiologic literature tends to focus on two types of biomarkers: 1) those that measure determinants of oxidative stress such as dietary pro- and antioxidants (biomarkers of exposure); and 2) those that measure products of ROS–induced

reactions (biomarkers of oxidation) or other events associated with oxidative stress (6). Only a limited number of biomarkers can measure several aspects of oxidative stress (e.g., oxidation of lipids, proteins and DNA), but these biomarkers are vulnerable to measurement error from non-oxidative stress processes (38, 46, 47). All of the biomarkers contribute to the understanding of oxidative stress, but since no single biomarker can reflect the oxidative stress phenomenon in its entirety (6, 111, 112), standard analytical methods limit a researcher to measuring only part of it.

Structural equation modeling (SEM) may overcome limitations of previous biomarker studies by offering a novel approach towards measuring and characterizing oxidative stress. SEM is a multivariate analytic technique that models a particular construct as an underlying latent (unobserved) variable derived from the shared covariance of multiple indicator (observed) variables (26, 27). The latent variable and indicator variables are modeled in a system of structural (causal) equations specified *a priori* (35), and the latent variable is identified from the modeled causal effects[1] it has on observed variables (26, 27). For example, the concept of oxidative stress can be modeled as a latent variable regressed on multiple biomarkers. If the biomarkers truly represent oxidative stress as an underlying causal construct, SEM should characterize it using estimated regression coefficients, called loading factors.

The overall goal of this study was to assess whether a latent variable can suitably identify and characterize oxidative stress. To achieve this goal, we addressed the following three specific research objectives. First, we assessed the content validity (26) of the latent variable by examining which markers had strong and weak oxidative stress loading factors. Next, we assessed construct validity (26) by investigating the associations of the latent variable with

[1] The authors would like to clarify the term "effects" in this context. As SEM is constructed under causal assumptions, the model estimated regression parameters are interpreted as effects (i.e. direct effect, indirect effect, total effect). The term does not mean that the SEM estimates reflect the causal effects that would have been observed under a well conducted randomized clinical trial, where the model assumptions would then differ. It is important to keep in mind that SEM estimated "effects" should not be interpreted as evidence of a causal link.

recognized pro-/antioxidant exposures. Finally, we compared those exposure associations to corresponding associations with the individual biomarkers. Assessing these differences allowed us to determine whether the latent variable provided new information, not offered by any individual biomarker.

To address the overall goal and each of the specific objectives, we used data from three cross-sectional studies that had collected demographic and behavioral data and measured plasma biomarkers of oxidative stress as well as biomarkers of relevant pro- and antioxidant exposures. The following four biomarkers of oxidation products and one biomarker of inflammation were selected *a priori* to identify and characterize oxidative stress: F_2-isoprostanes (FIP) – a marker of lipid peroxidation; fluorescent oxidation products (FOP) – a marker of non-specific oxidation (lipids, proteins, DNA); mitochondrial DNA copy number (MtDNA copies) – a marker cellular ROS-induced damage; γ-tocopherol (Gtoc) – a marker of metabolic response to oxidative stress; and C-reactive protein (CRP) – a marker of acute inflammation response. All five biomarkers were theorized to increase in response to oxidative stress.

MATERIALS and METHODS

Study Populations

Study of Race Stress and Hypertension (SRSH)

SRSH is a cross-sectional pilot study of 324 adult subjects across three different racial and ethnic groups: Non-Hispanic Whites (n = 124), African Americans (n = 99), and West African immigrants (n = 101). The West Africans (WA) were recruited in 2011 from multiple Atlanta, GA churches with ties to the West African immigrant population; the actual recruitment took place after church services or during a church festival. Non-Hispanic Whites (NHW) and African-Americans (AA) study subjects of similar age range were selected from among participants in the previously completed the Georgia Cancer Study (GCS) conducted at Emory

University. The GCS was a feasibility study in which approximately 800 adult participants were recruited at health fairs and similar events across the state of Georgia in 2007 - 2008. The recruitment and data collection for the West African immigrants and for GCS participants followed the same protocol. Eligibility for both studies included being a resident of the State of Georgia and age at recruitment between 25 and 74 years.

Markers of Adenomatous Polyps (MAP) study I and II

The Markers of Adenomatous Polyps I and II (MAP I and II) cross-sectional studies were conducted by the same principal investigator (RMB) using almost identical protocols. The methods for the two studies are described in detail elsewhere (79-81). Participants for MAP I (n = 474) were recruited from community gastroenterology practices in Winston-Salem and Charlotte, North Carolina, while MAP II (n = 233) was conducted among patients who received care at the Consultants in Gastroenterology, PA, a large, private practice clinic in Columbia, South Carolina. Eligible subjects were persons 30 - 74 years of age with no prior history of colorectal neoplasms who were scheduled to undergo elective colonoscopy. The study eligibility criteria were identical in both MAP studies. Cases (total n = 233) were subjects who were diagnosed with an incident colon or rectal adenoma at the time of their colonoscopy procedure. Controls (total n = 312) were all subjects who were free from all types of polyps during colonoscopy.

Data Collection

Questionnaire data

For the SRSH study, all eligible subjects filled out a questionnaire at the time of data collection immediately following informed consent. For the MAP studies, the questionnaires were mailed to eligible subjects prior to the colonoscopy procedure. Both questionnaires collected similar information on demographic characteristics, medical history, use of medications

and personal habits. Unlike SRSH the MAP studies also included a 153-item Willet Food Frequency Questionnaire to collect information on diet and nutritional supplement use.

Biomarker data

The blood collection protocol was similar in the SRSH and MAP studies. Blood was drawn into a pre-chilled Vacutainer tube then plunged into ice and covered to protect against light exposure. The samples were immediately transported to the study lab where they were immediately spun in a refrigerated centrifuge. Plasma and serum were separated and aliquotted into cryovials, and sealed with o-ring caps after inert gas displaced any oxygen in the vial. Samples were then stored at -80°C.

All biomarker analyses were performed by the Molecular Epidemiology and Biomarker Research Laboratory at the University of Minnesota (Minneapolis, MN). Plasma levels of α-carotene, β-carotene, α-tocopherol, γ-tocopherol, β-cryptoxanthin, lycopene, and lutein were assessed by high-performance liquid chromatography (113).

Plasma FIP concentrations were measured using a gas chromatography-mass spectrometry method (114). The FIP were extracted from the participants' samples using deuterium (4)-labeled 8-iso-prostaglandin F_2 alpha as an internal standard with unlabeled, purified F_2-isoprostane as a calibration standard. FOP were measured in plasma samples by the modified Shimasaki method as previously described (39, 115). A mixed solution of plasma and ethanol/ether was centrifuged for 10 minutes at 3,000 rpm after which 1.0 ml of supernatant was added to cuvettes for spectrofluorometric readings. Relative fluorescence intensity in units per milliliter of plasma at 360/430nm wavelengths was calculated by a spectrofluorometer. Quinine sulfate diluted in 0.1 N H_2SO_4 was used for calibration. Approximately 22% of the samples were serum rather than plasma due to a limited remaining supply of plasma samples from the study population. This was not expected to affect the FOP measurement or analysis as there was a high

level of correlation between a random sample of subjects on whom we measured FOP in both serum and plasma ($r = 0.9$; $p < 0.001$). High sensitive CRP was measured using latex-enhanced immunonephelometry on a Behring nephelometer II (BN-II) analyzer (inter-assay CV 4%; Behring Diagnostics, San Jose, CA). MtDNA copy number was analyzed in 2013 using real-time quantitative polymerase chain reaction (qPCR) as described previously (55). For the qPCR method we used two primers, one for MtDNA and one for nuclear DNA. The ratio of MtDNA to nuclear DNA was determined using serial dilution of a healthy referent genomic DNA sample.

Structural Equation Model and Statistical Methods

All SEMs were analyzed with Mplus software version 7.1, a statistical package developed specifically for these types of analyses (83). All SEM analyses originated from the baseline model where we theorized one continuous latent factor (here onward termed oxidative stress) composed of five continuous indicator variables chosen *a priori*: FIP, FOP, CRP, MtDNA copy number, and Gtoc. We standardized oxidative stress (mean = 0, variance = 1) to estimate loading factors for all five biomarkers. To aid empirical identification, the continuous indicator variables for oxidative stress were mathematically transformed to have similar variances; this transformation did not affect standardized results. All models were linear regressions with maximum likelihood estimation. A two-sided 0.05 p-value indicated statistical significance.

Before pooling the study populations, we first tested whether oxidative stress differed between the three race/ethnicity groups (NHW, AA, WA) in the SRSH dataset (Model 1), and then tested whether it differed between the three individual datasets (MAP I and II, and SRSH) (Model 2). This was done to assure that the latent variable was constructed the same way across groups; this is referred to as invariance testing. Mathematically, this can be expressed as whether the expected value for the indicator variables (y), conditional on the latent factor (η), are equal when group variables (g) are ignored or not ignored (116, 117).

$$E(y|\eta, g) = E(y|\eta) \quad (1)$$

If equation 1 holds, then the indicator variables are said to be invariant to group g, and a researcher can validly use the same metric and scale of the latent factor across the groups. We tested invariance within SRSH, and between MAPs and SRSH, to determine whether the biomarkers were indicators of the same concept across groups and to assess the feasibility of data pooling.

Invariance testing is done by setting unstandardized parameters equal to each other across groups and comparing nested models with respect to χ^2 tests and according to other common SEM fit statistics, such as the root mean squared error of approximation (RMSEA), the comparative fit index (CFI), the Tucker-Lewis index (TLI), and the standardized root mean square residual (SRMR). Equality constraints follow a hierarchy of weak factorial, strong factorial and strict factorial invariance (118). Each respective level increases the number equalities applied across groups. Weak factorial invariance assumes equality of biomarker loading factors, while strong factorial invariance additionally assumes equality of biomarker intercepts. Strict factorial invariance assumes invariance for the biomarker residual covariances. It is argued that weak factorial invariance is required for any group comparison, while strong factorial invariance is required for group comparisons among means (118). The exception is partial factorial invariance where only some biomarkers differ across groups (119). In this case a researcher has to make a qualitative determination of group equality.

Invariance testing was based on the following models and data sources.

Model 1: Stratifying by the three race/ethnicity groups (NHW, AA, WA), oxidative stress was modeled in relation to age (continuous, years), sex (binary), and body mass index (BMI) (<25, 25-<29.9, and ≥30 kg/m^2). We analyzed only SRSH data (n = 286).

Model 2: Stratifying by the three study populations (MAP I and II, and SRSH), oxidative stress was modeled in relation to age, sex, and BMI as predictor variables in the same fashion as in Model 1. The analytic dataset for this model included NHW non-cases from the MAP studies and all subjects from SRSH (n = 455).

For Models 1 and 2, weak and strong factorial invariance were tested first. If those tests were statistically significant, partial invariance was tested setting individual biomarker equalities. In partial invariance testing, we employed a "bottom-up" strategy starting with full inequality and then progressively added model restrictions. When we tested individual equalities the loading factor was set to be equal across all three groups with 2 degrees of freedom (df). If that restriction failed a test for statistical significance, the equality was set equal across two of the three groups (df = 1). Throughout the analyses we used the nested χ^2 test, with consideration for overall model fit statistics (RMSEA, CFI, TLI, SRMR). We reported both the unstandardized and the standardized loading factors (denoted λ_u and λ_s, respectively) for each biomarker along with their 95% confidence intervals (CI). The standardized loading factors were standardized to oxidative stress and each biomarker respectively. They are interpreted as a λ_s standard deviation (s.d.) change in the biomarker per one s.d. change in the latent oxidative stress variable.

After determining invariance, we evaluated content and construct validity of the latent oxidative stress variable. Content validity was deemed sufficient if more than one biomarker had moderate to strong loading factors, thereby indicating oxidative stress as their latent common cause. For construct validity, specific exposures (increased age, higher BMI, smoking, and drinking) were assumed to exert positive effects on oxidative stress, while others (aspirin use, α-carotene, β-carotene, α-tocopherol, β-cryptoxanthin, lycopene, and lutein) were expected to have negative effects. We then qualitatively compared the theorized and estimated directions of all moderate or strong exposure effects. Construct validity was deemed sufficient if most variables predicted oxidative stress in the theorized directions. The behavioral and demographic exposures

were analyzed in Model 3, and the plasma antioxidant exposures were analyzed in Model 4. These models were analyzed separately because some of the plasma measures were thought to be on the causal pathway between behavioral variables and oxidative stress. For plasma exposure variables (β_s), we reported the estimated model effects standardized to oxidative stress and each measured variable respectively. The effects of exposures measured by plasma markers are interpreted as β_s s.d. change in latent oxidative stress per one s.d. change in the respective exposure variable. For behavioral and demographic exposures we reported the effects standardized only to oxidative stress (β_s). These are interpreted as β_s s.d. change in oxidative stress per one unit change in the respective exposure.

For content and construct validity the following models and datasets were analyzed.

Model 3: Oxidative stress was modeled with the following variables as independent exposures: age (continuous, years), sex (binary), BMI (< 25, 25 - 29.9, and ≥ 30 kg/m^2), smoking history (never, former, current), drinking history (never, former, current), aspirin use (binary), non-steroidal anti-inflammatory drug use (NSAID) (binary) (Figure 2.1). We pooled data on the NHW non-cases from both MAP studies and on all subjects from SRSH (n = 413).

Model 4: Oxidative stress was modeled with the following continuous plasma exposures: α-carotene, β-carotene, α-tocopherol, β-cryptoxanthin, lycopene, and lutein (Figure 2.2). The model adjusted for all exposure variables in Model 3. As in model 3, data on the NHW non-cases from both MAP studies and on all subjects from SRSH (n = 382) were pooled.

In the final model we investigated whether the latent oxidative stress variable provided extra information beyond what can be obtained from evaluating individual biomarkers. This was achieved by comparing the model estimated effects of age, sex, BMI, smoking, drinking, and

aspirin and NSAID use on oxidative stress (Model 3) to the corresponding associations with each of the five individual markers: FIP, FOP, CRP, MtDNA copy number, and Gtoc (Model 5). A simple mean of the five associations for each exposure was calculated and reported.

Model 5: All five biomarkers were individually modeled on the behavioral and demographic exposures in Model 3: age, sex, BMI, smoking history, drinking history, aspirin use, and NSAID use (n = 413).

RESULTS

The SRSH population had a total of 324 subjects (NHW = 124, AA = 99, WA = 101). The average age of the population was 47.2 years with the WA participants on average five years younger than the NHW or AA participants. The proportion of women was the highest in the NHW subgroup (69.4%), followed by the WA (59.1%) and the AA (48.9%) subgroups. The proportions of obese (BMI $\geq$ 30), overweight (BMI 25 – 29.9), and normal weight (BMI < 25) did not differ greatly by race; 39.4% of all subjects were obese and 38.1% were overweight. Smoking and drinking history differed among the three groups with WA subjects having the highest proportions of never smokers and never drinkers. Plasma biomarkers of α-carotene, β-carotene, and lycopene were substantially higher among WA study subjects compared to those in the other two subgroups (Table 2.1).

As shown in Table 2.2, the pooled datasets from MAPs and SRSH included 482 NHW subjects. There was little difference in the average age between the MAP I and II study populations (both 56 years), but the SRSH participants were on average approximately 8 years younger. MAP II had the lowest proportion of women at 50.7%, while women comprised 66.0% and 69.4% of the MAP I and SRSH participants, respectively. The proportion of obese subjects was 34.8% in the pooled study population and did not vary meaningfully by study population. However, the proportion of normal weight and overweight subjects did differ by study

population, with MAP I and SRSH having the highest (42.1%) and the lowest (23.1%) proportions of normal weight persons, respectively. Smoking and drinking behaviors differed by study population. SRSH had no current smokers, but had the highest proportion of current drinkers (69.3%). MAP I had the lowest proportion of current drinkers (40.9%) but the highest proportion of current smokers (22.3%).

The SRSH subjects were racially and ethnically diverse, so we tested oxidative stress invariance across racial/ethnic subgroups in that study before pooling the data. Across the three racial/ethnic groups, the oxidative stress factor loadings for Gtoc and MtDNA copy number were invariant. Additionally, the factor loadings for FOP and CRP were invariant across NHW and AA, while FIP was invariant across AA and WA subjects (Table 2.3 – Model 1). Among NHW, oxidative stress was characterized by positive, statistically significant loading factors for FIP (λ_s = 0.56; 95% CI: 0.31 - 0.81), CRP (λ_s = 0.43; 95% CI: 0.25 - 0.61), and Gtoc (λ_s = 0.31; 95% CI: 0.16 - 0.46); the loading factor for MtDNA copy number was positive but not statistically significant (λ_s = 0.11; 95% CI: -0.11 - 0.32), and the corresponding estimate for FOP was negative and not statistically significant (λ_s = -0.07; 95% CI: -0.20 - 0.05). The FIP loading factor among AA, compared to among NHW, was weaker and not statistically significant (λ_s = 0.17; 95% CI: -0.09 - 0.43), while the λ_s for CRP was stronger (0.59; 95% CI: 0.42 - 0.76); the results for MtDNA copy number, Gtoc and FOP were similar. Oxidative stress loading factors differed markedly in WA subjects compared to those in NHW and AA. The most meaningful changes were for FOP, for which there was a statistically significant positive factor loading (λ_s = 0.56; 95% CI: 0.15 - 0.98), and for CRP, for which there was a negative, non-statistically significant factor loading (λ_s = -0.14; 95% CI: -0.58 - 0.30). Overall, we observed partial invariance between NHW and AA for four of the five biomarkers, and partial invariance between AA and WA, for three of the five biomarkers. However, as FIP was not invariant between NHW and AA, we restricted the other models to NHW.

Before pooling the data from the SRSH and the two MAP studies we tested oxidative stress invariance across the three study populations. We found that MtDNA copy number, Gtoc, and FOP loading factors were invariant across all three studies, CRP was invariant between MAP I and SRSH, and FIP was invariant between SRSH and MAP II (Table 2.4 – Model 2). Despite the differences in the biomarker loading factors, oxidative stress was similarly characterized in all three populations. FIP had the strongest positive loading, followed by CRP and Gtoc, which had statistically significant but more moderate factor loadings; MtDNA copy number and FOP were null. The partial invariance of loading factors was sufficient to warrant testing the invariance of the model intercepts. There was intercept invariance for Gtoc and FOP across all three studies, and FIP invariance between MAP I and II. The intercept for CRP was borderline statistically significant across all three studies ($p = 0.05$; data not shown). Given the partial loading factor invariance and the partial intercept invariance in Model 2, the pooling of data on NHW participants from all three studies was deemed feasible.

Using the pooled data we assessed content validity of oxidative stress from the biomarker factor loadings, and construct validity from the estimated exposure effects on latent oxidative stress. As in Model 2, oxidative stress was characterized by positive, statistically significant factor loadings from FIP ($\lambda_s = 0.66$; 95% CI: 0.53 - 0.78), Gtoc ($\lambda_s = 0.51$; 95% CI: 0.38 - 0.63), and CRP ($\lambda_s = 0.44$; 95% CI: 0.32 - 0.56). Different from that in Model 2, the loading factor for MtDNA copy number was negative and weak, but borderline statistically significant ($\lambda_s = -0.18$; 95% CI: -0.35 - ¯0.01) (Table 2.5 – Model 3).

We assessed construct validity of oxidative stress from behavioral and demographic exposure variables (Model 3) and from plasma antioxidant levels (Model 4). In Model 3, the associations of BMI, a history of smoking, and regular aspirin use with oxidative stress were statistically significant and in the theorized directions; the associations of higher age and regular NSAID use with oxidative stress were null (Table 2.5). Compared to normal weight individuals,

obesity (BMI ≥ 30) had the strongest model effect on oxidative stress among all predictors (β_s = 1.11; 95% CI: 0.80 - 1.42), with overweight having a slightly less pronounced effect (β_s = 0.56; 95% CI: 0.26 - 0.87) with p for trend < 0.01. For smoking history there was also a statistically significant trend in the theorized direction, with β_s estimates of 0.16 (95% CI: -0.13 - 0.45) for former smoking and 0.40 (95% CI: 0.02 - 0.78) for current smoking (p for trend = 0.02). Regular aspirin use had a negative model effect on oxidative stress (β_s = -0.39; 95% CI: ¯0.68 - ¯ 0.11). A history of drinking had a negative model effect on oxidative stress (β_s = -0.24; 95% CI: -0.66 - 0.18, for former drinking; β_s = -0.46; 95% CI: ¯0.79 - ¯0.14, for current drinking). In Model 4 we included plasma antioxidant exposures. Higher levels of β-carotene, α-carotene and α-tocopherol were inversely associated with oxidative stress, although the estimates for the latter two were not statistically significant. Higher levels of lycopene were positively associated with oxidative stress (Table 2.5).

When the results from Model 3 were compared to the regression estimates for the individual biomarkers (Model 5) the effects of exposures on the latent oxidative stress variable differed considerably from those for the individual markers (Table 2.6). For almost all exposures other than age and NSAID use (for which the estimates were null) the β_s estimates in relation to the latent oxidative stress variable were stronger than any individual biomarker-specific coefficient; and all were much stronger than the simple mean (μ). For example, the SEM effect of obesity on oxidative stress was twice as strong as the mean of regression coefficients-specific estimates (β_s = 1.11 vs. 0.45 respectively). Additionally, the magnitudes and directions of the associations between an exposure with each biomarker varied considerably. For instance, the associations of aspirin use with FIP, Gtoc, CRP, FOP, and MtDNA were -0.29, -0.27, -0.03, 0.11, and 0.26, respectively. Likewise, gender was strongly associated with FIP (0.73), moderately associated with CRP (0.38), and weakly or not associated with Gtoc (0.17), FOP (0.16), and MtDNA (0.06).

DISCUSSION

In this study using SEM we identified a latent oxidative stress variable that was characterized primarily by three (FIP, CRP, and Gtoc) of five (FOP and MtDNA copy number) *a priori* selected biomarkers of oxidative stress. Plasma FIP is a well-known biomarker of oxidative stress, sometimes referred to as the "gold standard" indicator of lipid peroxidation (37, 40, 120). CRP, while recognized as a non-specific biomarker of inflammation, is correlated with FIP and other biomarkers of oxidative stress (43, 66-69). Oxidative stress is also closely linked with inflammation as it can increase many inflammatory transcription factors, such as nuclear factor kappa B (NF-κB) and interleukin-8 (IL-8) (65, 121). Chemically, Gtoc has antioxidant properties (57), but findings regarding Gtoc supplementation in relation to markers of oxidative stress have been mixed (58-61). It is possible that circulating levels of plasma Gtoc are more representative of its metabolism by cytochrome P450 than its intake levels. There is evidence that oxidative stress and inflammation may negatively alter the metabolism of Gtoc (57, 60), thereby increasing its levels in the circulation. Circulating levels of plasma Gtoc are positively associated with FIP and CRP (43), and Cooney et al. proposed that measured levels are more indicative of the presence of underlying conditions than of being a cause of those conditions (62). FOP has been proposed as a non-specific marker of lipid, protein, and DNA oxidation (46, 47). Because this sensitive, stable biomarker (39) has been linked to pro-oxidant behaviors (38), it was included as an oxidative stress indicator variable. MtDNA copy number is a marker of structural variation in MtDNA (122) and was included as an indicator variable because of the role of mitochondria in ROS production (123) and its association with thiobarbituric acid-reactive substances – an oxidative stress biomarker (56).

Based on the evidence about the biochemical processes influencing FIP, CRP, and Gtoc levels, and on the observed shared covariance of these markers in the SEM, we conclude that these three biomarkers share a common cause, which may be justifiably called "oxidative stress."

The roles of mtDNA copy number and FOP are unclear, and these two biomarkers contributed little to the characterization of the latent oxidative stress variable in the present study.

In the analyses to evaluate pro- and antioxidant exposures and their relation to oxidative stress, almost all of the estimated associations involving our primary exposures were either null or in the theorized direction. In particular, our analysis indicated that oxidative stress was positively associated with BMI and smoking, observations that are consistent with the prior evidence. Higher levels of obesity are known to be positively associated with oxidative stress biomarkers (107). Smoking is also known to increase oxidative stress (100), and has been positively associated with FIP, CRP, and Gtoc in previous studies (43, 62, 105, 124). According to the "oxidative stress theory", increased macromolecular oxidation by the ROS influences the rate of aging (125). Although the link between oxidative stress and aging is plausible, the observed associations between higher age and biomarkers of oxidation are inconsistent (43, 105, 126), and in the present study, age was not a predictor of the latent oxidative stress variable. As inflammation and ROS production are closely linked (65), aspirin use is expected to reduce oxidative stress (108, 127), and our findings are consistent with this.

A somewhat unexpected result was the inverse association between alcohol consumption and oxidative stress. Although this may have been a chance finding, the observed result for alcohol warrants further evaluation. While pure alcohol is pro-oxidant (101, 102) many alcoholic beverages contain polyphenols which could decrease oxidative stress or inflammation (103, 104). The opposite effects of pure alcohol and certain alcoholic beverages may explain the inconsistent literature comparing oxidative stress biomarkers in drinkers and non-drinkers (43, 128).

In the analyses involving plasma antioxidant levels, the results were less convincing than those for obesity, smoking, and aspirin. Although the nutrients assessed in the present study have antioxidant properties (129), the associations of the nutrients with biomarkers of oxidative stress

remain inconsistent (40, 88, 130). In the present study, α-carotene, β-carotene, and α-tocopherol were all inversely associated with oxidative stress, but only the β-carotene-oxidative stress association was at least moderately strong. The results for lutein and β-cryptoxantin were null, and contrary to expectation, the association of lycopene with the latent oxidative stress variable was positive. The result for lycopene is difficult to explain, but could be attributable to potential interactions with other antioxidants, smoking, or other variables that were not included in the models (45, 131).

Another important observation in the present study is the consistently stronger associations of pro- and antioxidant exposures with the latent oxidative stress variable relative to the corresponding associations with the individual biomarkers that comprise the latent variable. These results suggest that the latent variable provides unique information not attainable by any individual biomarker. Furthermore, the inconsistent magnitudes and directions of the associations between the exposures and the individual biomarkers may explain the inconsistent biomarker-specific associations in previous studies, and justify the use of SEM in future research.

We observed that the oxidative stress latent variable varied by race/ethnicity, and was associated with gender. FOP was the strongest positive loading factor for WA immigrants, but null for NHW or AA. FIP was the strongest loading factor for NHW but not for AA or WA. Gtoc was the only biomarker that was consistently linked to oxidative stress across the three racial/ethnic groups. Additionally, we observed that oxidative stress was higher among women than among men. Biologically, this is plausible since sex hormones may influence oxidative stress (132-134), but the epidemiologic evidence for this is mixed; FIP is likely higher in women (43, 105), but the findings for CRP in these regards are inconsistent (135-137). It appears that the observed differences and similarities in loading factors and predictors of oxidative stress across the sex and racial/ethnic groups are worth further study.

Perhaps the most important limitation of our study is the cross-sectional design. This design feature precludes determining the temporal relation between the determinants and markers of oxidative stress; although for the purposes of SEM it must be assumed that exposure precedes the outcome. Only a properly designed follow-up study can definitively ascertain the sequence of events to justify the cause-and-effect assumptions required for SEM-based analyses. Also, to obtain a sufficient population size, we pooled data from three studies. The invariance testing demonstrated that the model measured similar latent constructs among NHW subjects in all three studies; however, the results did differ across race/ethnicity groups, and thus our findings can only be generalized to NHW. We cannot be sure that the latent variable in this study reflects exclusively, or wholly, the construct of oxidative stress and that oxidative stress is completely characterized by the available biomarkers. Given the close interrelation of oxidative stress with inflammation and the importance of CRP in these data, it is also possible that the latent variable in this study represents inflammation or some combination of inflammation and oxidative stress. On the other hand, it is worth pointing out that regular NSAID use in this study was positively associated with CRP but not with the latent oxidative stress variable.

In conclusion, based on the prior evidence about the biochemical processes influencing FIP, CRP, and Gtoc levels, and on the observed shared covariance of these measures in the SEM, we conclude that these three biomarkers share a common cause, which may be justifiably called "oxidative stress." The roles of mtDNA copy number and FOP are unclear, and these two biomarkers appear to contribute little to the characterization of the latent oxidative stress variable. Our study illustrates how SEM methodology allows construction of a composite biomarker-based latent variable, and permits answering research questions that cannot be addressed by evaluating one biomarker at a time. The methods used in the present study can be applied to similar situations and in other areas of research in which it is necessary to combine different, but imperfect measurements to describe a complex biologic phenomenon.

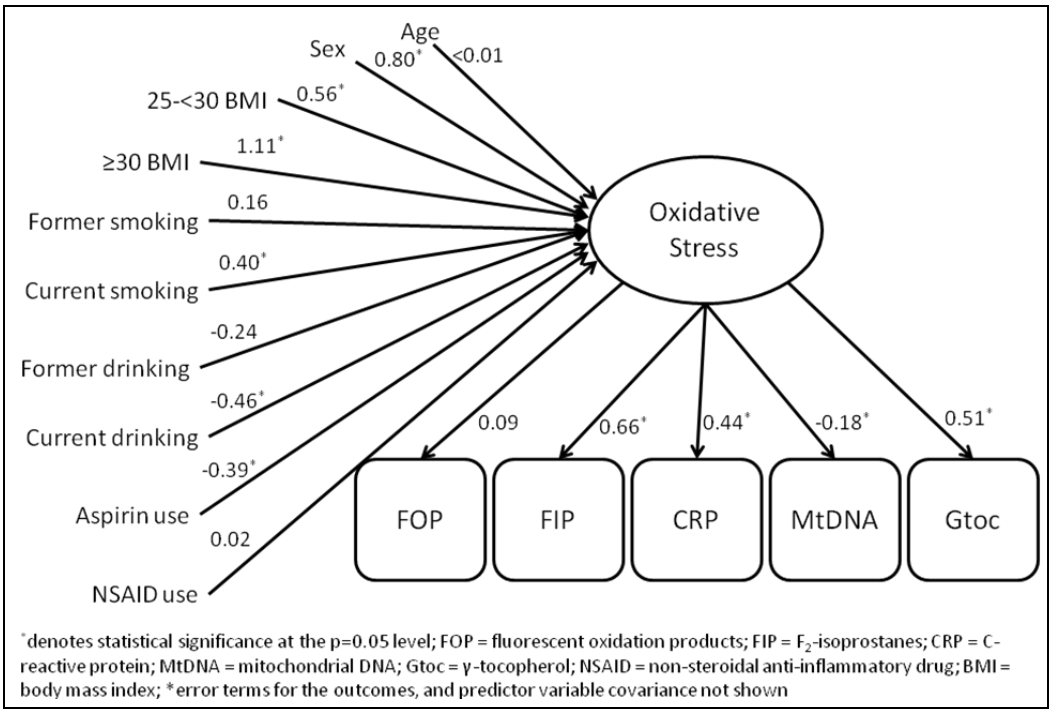

Figure 2.1. Structural equation model of primary exposures on oxidative stress in three pooled datasets

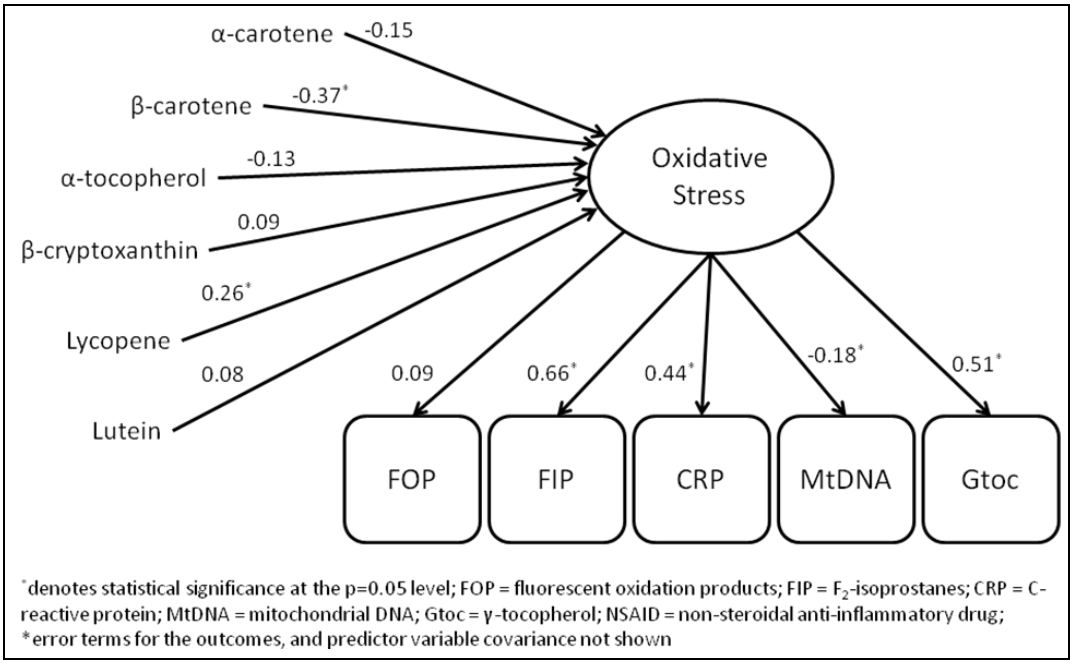

Figure 2.2. Structural equation model of secondary exposures on oxidative stress in three pooled datasets; not shown are age, sex, BMI, smoking, drinking, aspirin, NSAID covariates

Table 2.1. Selected baseline characteristics of subjects stratified by race in the Study of Race, Stress, and Hypertension

	NHW (n = 124)		AA (n = 99)		WA (n = 101)		Total (n = 324)	
Variable	μ or n	% or s.d.	μ or n	% or s.d.	μ or n	% or s.d.	μ or n	% or s.d.
Age, years	48.4	13.0	48.7	10.6	43.5	11.0	47.2	11.9
Sex (% women)	86	69.4	46	48.9	55	59.1	195	60.8
BMI, kg/m^2								
< 25	28	23.1	18	19.2	20	23.0	70	22.4
25 - 29.9	47	38.8	43	45.7	27	31.0	119	38.1
≥ 30	46	38.0	33	35.1	40	46.0	123	39.4
Smoker								
Never	53	59.6	40	28.6	88	92.6	188	75.8
Ever	36	40.5	16	71.4	7	7.4	60	24.2
Drinker								
Never	13	14.8	13	24.1	43	46.7	72	29.8
Former	14	15.9	22	40.7	28	30.4	65	26.9
Current	61	69.3	19	35.2	21	22.8	105	43.4
Aspirin use	68	76.4	18	32.7	19	19.6	59	23.7
NSAID use	21	23.6	15	27.3	15	15.5	68	27.3
Plasma biomarkers								
α-carotene, μg/dl	2.7	3.4	3.3	3.0	25.9	17.0	10.9	15.0
β-carotene, μg/dl	11.1	8.9	15.9	16.8	40.1	27.2	22.5	22.9
α-tocopherol, mg/dl	1.0	0.29	1.1	0.28	0.8	0.19	1.0	0.28
Lutein, μg/dl	16.0	7.8	25.8	12.6	23.6	9.5	21.1	10.5
Lycopene, μg/dl	33.4	13.8	35.8	15.4	65.9	27.0	45.3	24.8
β-cryptoxanthin,μg/dl	5.7	3.9	9.6	16.8	8.6	5.0	7.7	9.1
Ferritin, mg/dl	116.9	128.0	127	104.6	139.9	352.0	128.1	224.5
Cholesterol, mg/dl	193.4	39.0	216	46.4	188.8	38.6	197.9	42.4
FIP, pg/ml	82.5	41.0	51.1	20.1	33.8	9.0	56.7	34.6
FOP, avg. std. ref. adj.	0.04	0.02	0.04	0.01	0.05	0.02	0.04	0.02
CRP, μg/ml	3.9	8.2	3.6	4.4	2.7	5.7	3.4	6.5
MtDNA, rel. nucl.DNA	3.2	0.7	2.9	0.9	3.4	0.8	3.2	0.8
Gtoc, mg/dl	0.22	0.09	0.25	0.11	0.14	0.05	0.20	0.09

Symbols and abbreviations: NHW = non-Hispanic whites; AA = African Americans; WA = West African immigrants; μ = mean; s.d. = standard deviation; BMI = body mass index; NSAID = non-steroidal anti-inflammatory drug; FIP = F_2-isoprostanes; FOP = fluorescent oxidation products; CRP = C-reactive protein; MtDNA = mitochondrial DNA; Gtoc = γ-tocopherol; avg. std. ref. adj. = average standard reference adjusted; rel. nucl. DNA = MtDNA relative to nuclear DNA;

Table 2.2. Baseline characteristics of non-Hispanic white subjects in three datasets

	MAP I (n = 212)		MAP II (n = 146)		SRSH (n = 124)		Total (n = 482)	
Variable	μ or n	% or s.d.	μ or n	% or s.d.	μ or n	% or s.d.	μ or n	% or s.d.
Age	56.7	10.3	55.2	8.0	48.4	13.0	54.1	11.0
Sex (% women)	140	66.0	74	50.7	86	69.4	300	62.2
BMI, kg/m^2								
< 25	88	42.1	42	29.8	28	23.1	158	33.6
25 - 29.9	46	22.0	56	39.7	47	38.8	149	31.6
≥ 30	75	35.9	43	30.5	46	38.0	164	34.8
Smoker								
Never	86	42.6	66	46.5	53	59.6	205	47.3
Former	71	35.2	58	40.9	36	40.5	165	38.1
Current	45	22.3	18	12.7	0	0.0	63	14.6
Drinker								
Never	87	41.8	27	19.0	13	14.8	127	29.0
Former	36	17.3	29	20.4	14	15.9	79	18.0
Current	85	40.9	86	60.6	61	69.3	232	53.0
Aspirin use	65	31.0	56	39.7	68	76.4	142	32.3
NSAID use	65	31.0	52	36.9	21	23.6	151	34.3
Plasma biomarkers								
α-carotene, μg/dl	3.5	3.8	3.4	3.1	2.7	3.4	3.3	3.5
β-carotene, μg/dl	16.2	15.6	17.7	17.7	11.1	8.9	15.3	15.1
α-tocopherol, mg/dl	1.16	0.51	1.34	0.71	1.00	0.29	1.17	0.55
Lutein, μg/dl	16.9	8.6	15.6	6.7	16.0	7.8	16.3	7.8
Lycopene, μg/dl	26.7	13.9	26.0	11.0	33.4	13.8	28.3	13.4
β-cryptoxanthin	6.6	5.2	8.0	7.3	5.7	3.9	6.8	5.7
Cholesterol, mg/dl	207.7	38.3	203.7	42.4	193.4	39.0	202.7	40.1
FIP, pg/ml	87.4	36.6	82.1	33.4	82.5	41.0	84.7	37.0
FOP, avg. std. ref.adj.	0.05	0.10	0.04	0.01	0.04	0.02	0.04	0.06
CRP, μg/ml	5.6	6.3	4.1	5.4	3.9	8.2	4.5	5.6
MtDNA, rel.nucl.DNA	0.9	1.0	4.3	1.7	3.2	0.7	2.6	2.1
Gtoc, mg/dl	0.23	0.11	0.18	0.09	0.22	0.09	0.21	0.10

Symbols and abbreviations: MAP = Markers of Adenomatous Polyps; SRSH = Study of Race, Stress, Hypertension; μ = mean; s.d. = standard deviation; BMI = body mass index; NSAID = non-steroidal anti-inflammatory drug; FIP = F_2-isoprostanes; FOP = fluorescent oxidation products; CRP = C-reactive protein; MtDNA = mitochondrial DNA; Gtoc = γ-tocopherol; avg. std. ref. adj. = average standard reference adjusted; rel. nucl. DNA = MtDNA relative to nuclear DNA;

Table 2.3. Test for oxidative stress weak factorial invariance stratified by race in the Study of Race, Stress, and Hypertension

No weak factorial invariance

	Among Non-Hispanic White (n=121)				**Among African American (n = 93)**				**Among West Africans (n = 101)**					
	λ_u	λ_s	95% LCI	95% UCI	λ_u	λ_s	95% LCI	95% UCI	λ_u	λ_s	95% LCI	95% UCI	Model Fit Statistics	
MtDNA	-0.18	-0.05	-0.52	0.43	0.23	0.13	-0.15	0.42	0.70	0.13	-0.20	0.46	$\chi^{2\ pvalue}$	0.05
FIP	3.23	0.60	0.36	0.84	0.24	0.24	-0.09	0.57	0.10	0.06	-0.31	0.44	RMSEA	0.06
FOP	-0.62	-0.13	-0.37	0.10	-0.13	-0.10	-0.41	0.21	3.39	0.61	0.21	1.01	CFI	0.76
CRP	1.29	0.44	0.22	0.65	0.62	0.55	0.17	0.93	-0.44	-0.12	-0.50	0.26	TLI	0.66
Gtoc	1.65	0.39	0.17	0.61	0.55	0.28	0.01	0.55	1.21	0.47	0.14	0.80	SRMR	0.10

Partial weak factorial invariance

	λ_u	λ_s	95% LCI	95% UCI	λ_u	λ_s	95% LCI	95% UCI	λ_u	λ_s	95% LCI	95% UCI		
MtDNA	0.35	0.11	-0.11	0.32	0.35	0.12	-0.12	0.36	0.35	0.07	-0.07	0.21	$\chi^{2\ pvalue}$	0.10
FIP	2.66	0.56	0.31	0.81	0.28	0.17	-0.09	0.43	0.28	0.18	-0.08	0.44	RMSEA	0.05
FOP	-0.30	-0.07	-0.20	0.05	-0.30	-0.14	-0.37	0.09	3.04	0.56	0.15	0.98	CFI	0.82
CRP	1.12	0.43	0.25	0.61	1.12	0.59	0.42	0.76	-0.51	-0.14	-0.58	0.30	TLI	0.77
Gtoc	1.16	0.31	0.16	0.46	1.16	0.34	0.17	0.51	1.16	0.46	0.20	0.72	SRMR	0.10

Symbols and abbreviations: λ_u = unstandardized loading factor; λ_s = standardized loading factor; LCI = lower confidence interval; UCI = upper confidence interval; χ^2 = Chi-squared; RMSEA = root mean squared error of approximation; CFI = comparative fit index; TLI = Tucker-Lewis index; SRMR = standardized root mean square residual; FIP = F_2-isoprostanes; FOP = fluorescent oxidation products; CRP = C-reactive protein; MtDNA = mitochondrial DNA; Gtoc = γ-tocopherol;

Table 2.4. Test for oxidative stress weak factorial invariance stratified by study population in Model 2

No invariance

	Among MAP II (n = 136)				Among MAP I (n = 201)				Among SRSH (n = 118)					
	λ_u	λ_s	95% LCI	95% UCI	λ_u	λ_s	95% LCI	95% UCI	λ_u	λ_s	95% LCI	95% UCI	Model Fit Statistics	
MtDNA	-0.67	-0.16	-0.41	0.08	0.07	0.05	-0.17	0.26	0.06	0.05	-0.43	0.52	$\chi^{2\ pvalue}$	0.30
FIP	4.43	0.93	0.75	1.12	1.93	0.60	0.44	0.76	-2.48	-0.60	-0.84	-0.36	RMSEA	0.02
FOP	0.10	0.12	-0.11	0.35	0.50	0.11	-0.08	0.30	0.14	0.13	-0.10	0.37	CFI	0.97
CRP	3.06	0.61	0.44	0.78	1.56	0.41	0.27	0.56	-1.29	-0.44	-0.65	-0.22	TLI	0.96
Gtoc	2.42	0.51	0.34	0.68	1.64	0.43	0.26	0.60	-1.28	-0.39	-0.61	-0.17	SRMR	0.06

Partial invariance

	Among MAP II (n = 136)				Among MAP I (n = 201)				Among SRSH (n = 118)					
	λ_u	λ_s	95% LCI	95% UCI	λ_u	λ_s	95% LCI	95% UCI	λ_u	λ_s	95% LCI	95% UCI		
MtDNA	-0.03	-0.01	-0.09	0.07	-0.03	-0.02	-0.20	0.17	-0.03	-0.02	-0.23	0.20	$\chi^{2\ pvalue}$	0.24
FIP	3.74	0.86	0.66	1.06	2.00	0.61	0.45	0.76	3.74	0.75	0.59	0.90	RMSEA	0.03
FOP	0.03	0.04	-0.15	0.23	0.03	0.01	-0.03	0.04	0.03	0.03	-0.10	0.15	CFI	0.96
CRP	2.96	0.61	0.43	0.80	1.55	0.40	0.27	0.53	1.55	0.45	0.30	0.60	TLI	0.94
Gtoc	1.78	0.40	0.27	0.53	1.78	0.46	0.34	0.57	1.78	0.45	0.32	0.59	SRMR	0.07

Symbols and abbreviations: MAP = Markers of Adenomatous Polyps; SRSH = Study of Race, Stress, and Hypertension; λ_u = unstandardized loading factor; λ_s = standardized loading factor; LCI = lower confidence interval; UCI = upper confidence interval; χ^2 = Chi-squared; RMSEA = root mean squared error of approximation; CFI = comparative fit index; TLI = Tucker-Lewis index; SRMR = standardized root mean square residual; FIP = F_2-isoprostanes; FOP = fluorescent oxidation products; CRP = C-reactive protein; MtDNA = mitochondrial DNA; Gtoc = γ-tocopherol;

Table 2.5. Estimated model effects of exposure variables on latent oxidative stress, and loading factors for oxidative stress in three pooled datasets

Model estimated effects on oxidative stress

	β_s	95% LCI	95% UCI	p-value	Model Fit Statistics	
Age	< 0.01	-0.01	0.01	0.93	$\chi^{2\ pvalue}$	< 0.01
Sex (vs. Men)	0.80	0.53	1.08	< 0.01	RMSEA	0.05
BMI, kg/m^2					CFI	0.73
< 25	referent				TLI	0.64
25 - 29.9	0.56	0.26	0.87	< 0.01	SRMR	0.04
≥ 30	1.11	0.80	1.42	< 0.01		
Smoking						
Never	referent					
Former	0.16	-0.13	0.45	0.27		
Current	0.40	0.02	0.78	0.04		
Drinking						
Never	referent					
Former	-0.24	-0.66	0.18	0.26		
Current	-0.46	-0.79	-0.14	0.01		
Aspirin use	-0.39	-0.68	-0.11	0.01		
NSAID use	0.02	-0.26	0.29	0.91		
Plasma markers						
α-carotene	-0.15	-0.31	0.01	0.06		
β-carotene	-0.37	-0.53	-0.21	<0.01		
α-tocopherol	-0.13	-0.27	0.01	0.06		
Lutein	0.08	-0.06	0.23	0.25		
Lycopene	0.26	0.14	0.39	<0.01		
β-cryptoxanthin	0.09	-0.05	0.22	0.22		

Loading factors for Oxidative Stress

	λ_s	95% LCI	95% UCI	p-value
MtDNA	-0.18	-0.35	-0.01	0.04
FIP	0.66	0.53	0.78	< 0.01
FOP	0.09	-0.05	0.23	0.21
CRP	0.44	0.32	0.56	< 0.01
Gtoc	0.51	0.38	0.63	< 0.01

Symbols and abbreviations: λ_s = standardized loading factor; β_s = standardized model effect; LCI = lower confidence interval; UCI = upper confidence interval; χ^2 = Chi-squared; RMSEA = root mean squared error of approximation; CFI = comparative fit index; TLI = Tucker-Lewis index; SRMR = standardized root mean square residual; FIP = F_2-isoprostanes; FOP = fluorescent oxidation products; CRP = C-reactive protein; MtDNA = mitochondrial DNA; Gtoc = γ-tocopherol; BMI = body mass index; NSAID = non-steroidal anti-inflammatory drug;

Table 2.6. Estimated regression parameters of exposures on individual oxidative stress biomarkers in three pooled datasets

	FIP	**Gtoc**	**CRP**	**FOP**	**MtDNA**	
Exposure	β_{FIP}	β_{Gtoc}	β_{CRP}	β_{FOP}	β_{MtDNA}	μ_{all}
Age	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
Sex (vs. Men)	0.73**	0.17	0.38**	0.16	0.06	0.30
BMI, kg/m^2						
< 25			referent			
25 - 29.9	0.31*	0.33**	0.40**	-0.05	0.32*	0.26
≥ 30	0.79**	0.58**	0.56**	0.09	0.25	0.45
Smoking						
Never			referent			
Former	0.13	0.19	-0.10	-0.03	-0.13	0.01
Current	0.24	0.12	0.19	0.29	-0.38*	0.09
Drinking						
Never			referent			
Former	0.17	-0.55**	0.04	-0.10	0.42*	< 0.01
Current	-0.07	-0.46**	-0.22	0.07	0.42**	-0.05
Aspirin use	-0.29**	-0.27**	-0.03	0.11	0.26**	-0.04
NSAID use	-0.10	-0.10	0.26*	0.07	0.13	0.05

Symbols and abbreviations: β_{xxx} = standardized regression coefficient for biomarker; μ_{all} = simple mean of all biomarkers; FIP = F_2-isoprostanes; FOP = fluorescent oxidation products; CRP = C-reactive protein; MtDNA = mitochondrial DNA; Gtoc = γ-tocopherol; BMI = body mass index; NSAID = non-steroidal anti-inflammatory drug; *significance at the p=0.05 level; **significance at the p=0.05 level

CHAPTER 3. A NEW LOOK AT OXIDATIVE STRESS AND ITS ASSOCIATION WITH COLORECTAL ADENOMA: A LATENT VARIABLE, STRUCTURAL EQUATION MODELING APPROACH

ABSTRACT

Background: Oxidative stress has been implicated in many adverse health conditions including colorectal adenoma and carcinoma. This complex multifaceted phenomenon is usually measured by imperfect biomarkers resulting in weak and inconsistent associations with health outcomes. The analytic technique of structural equation modeling (SEM) can reduce measurement error of individual biomarkers through a latent construct and may demonstrate stronger associations.

Methods: Using a pooled dataset of two previously conducted cross-sectional studies (Markers of Adenomatous Polyps I and II; n = 526), we modeled a latent oxidative stress variable from five biomarkers: F_2-isoprostanes (FIP), fluorescent oxidation products (FOP), mitochondrial DNA copy number (MtDNA), γ-tocopherol (Gtoc), and C-reactive protein (CRP). Using SEM, we modeled a simplified causal pathway of pro-/antioxidant exposures, oxidative stress, and colorectal adenoma.

Results: The latent oxidative stress variable was strongly and positively associated with colorectal adenoma (odds ratio = 2.61, 95% confidence interval: 1.25-5.46). Oxidative stress was characterized by positive loading factors from FIP, Gtoc, and CRP. Additionally, BMI and other exposure variables were associated with colorectal adenoma in a pathway mediated by oxidative stress.

Conclusion: A better measurement of oxidative stress can be attained through SEM with latent variables. Oxidative stress is positively associated with colorectal adenoma, and exposure variables are likely to be associated with adenoma through this mechanism. The SEM methodology can be applied to studying other oxidative stress-related health outcomes. It can be

extended to other areas of research where it is necessary to combine different, but imperfect measurements to describe a complex biologic phenomenon.

INTRODUCTION

Oxidative stress has long been defined as an imbalance between exogenous and endogenously produced pro- and antioxidants, resulting in excess reactive oxygen species (ROS) (5). Increased ROS, produced from the chemical and biological processes of oxidative stress, have been implicated in many adverse effects on human health (2), including the initiation, promotion, and progression of carcinogenesis (3, 4, 65, 93, 94, 138). The basic science evidence that has given rise to the oxidative stress-induced neoplasia hypothesis, has not been consistently confirmed in the epidemiologic literature (90, 95-99). Furthermore, the attempts to prevent oxidative stress-induced cancer through supplementary antioxidant interventions in human clinical trials have yielded null results, or sometimes even demonstrated harmful effects (9). The inconsistent results across basic science, epidemiologic, and clinical trials research points to a lack of knowledge on how to influence or measure oxidative stress in humans.

Measurement of oxidative stress is normally done through biomarkers, as this complex phenomenon is not directly observable *in vivo* (36). There are numerous biomarkers of oxidative stress, many of which reflect a specific biologic or chemical aspect of this process (6). As oxidative stress is multifaceted (139), reliance on any single biomarker may be inadequate (111, 112). Additionally, there are numerous determinants of oxidative stress, but few have shown consistent results with biomarkers or health outcomes (38, 40, 43, 90, 95, 140). Because of the inconsistency, researchers investigating the burden of oxidative stress exposures have proposed to combine oxidative stress measurements, and pro-/antioxidant behavioral characteristics, into a score variable (10, 12-14, 141). Using this approach, studies have demonstrated that an oxidative balance score (OBS) had stronger associations with colorectal neoplasia than the corresponding associations of the individual variables when considered alone (16-18, 142).

Although studies have found stronger associations between health outcomes and the OBS, it remains unclear whether the score truly captures oxidative stress-related exposures, or is just a measure of general good health. It is possible that individual antioxidant nutrients can influence carcinogenesis through mechanisms other than oxidative stress (19-21), and it is a certainty that behavioral variables such as smoking, drinking, and medication use may directly affect many carcinogenic pathways (22-25). A recent study by Kong et al. found that, in addition to colorectal adenoma, OBS was associated with plasma F_2-isoprostanes (FIP), fluorescent oxidative products (FOP), and C-reactive protein (CRP); however only the results for FIP and CRP were in the theorized directions (18). The unexpected result of FOP confirms that further study of oxidative stress, its determinants and biomarkers, and colorectal adenoma is warranted.

In a previous study, we used the analytic method of structural equation modeling (SEM) to obtained a measurement of oxidative stress using a latent (unobserved) variable, composed of various biomarkers of oxidation (143). SEM is a multivariate technique that models theoretical structural (causal) pathways, specified *a priori*, in observational data (26, 27). The method frequently incorporates the use of latent variables as measurements of unobservable constructs. For instance, a researcher could construct a theoretical model where specific pro-/antioxidant exposures have direct effects[2] on oxidative stress, which in turn could have direct effects on colorectal adenoma. The two direct effects would make up one indirect effect from exposures to adenoma mediated by oxidative stress. The advantage of SEM is its ability to reduce measurement error through latent variables, and simultaneously estimate multiple parameters through complex likelihoods (26, 27).

[2] The authors would like to clarify the term "effects" in this context. As SEM is constructed under causal assumptions, the model estimated regression parameters are interpreted as effects (i.e. direct effect, indirect effect, total effect). The term does not mean that the SEM estimates reflect the causal effects that would have been observed under a well conducted randomized clinical trial, where the model assumptions would then differ. It is important to keep in mind that SEM estimated "effects" should not be interpreted as evidence of a causal link.

In this study, we assessed whether a latent variable of oxidative stress is associated with newly diagnosed colorectal adenoma. We were also interested in two secondary questions: 1) How does latent oxidative stress variable in this analysis compare to our previous study (143), which used a somewhat different dataset and did not include adenoma as the outcome of interest; and 2) Do pro-/antioxidant exposures have indirect associations with colorectal adenoma, mediated by oxidative stress. To address these questions, we used pooled data from two cross-sectional colorectal adenoma studies which collected demographic and behavioral data, as well as plasma biomarkers of oxidative stress. The following four biomarkers of oxidation products and one biomarker of inflammation were selected *a priori* to identify and characterize oxidative stress: F_2-isoprostanes (FIP) – a marker of lipid peroxidation; fluorescent oxidation products (FOP) – a marker of non-specific oxidation (lipids, proteins, DNA); mitochondrial DNA copy number (MtDNA) – a marker of cellular damage; γ-tocopherol (Gtoc) – a marker of metabolic response to oxidative stress; and C-reactive protein (CRP) – a marker of acute inflammation response. All five biomarkers were theorized to increase in response to oxidative stress.

MATERIALS and METHODS

Study Population

Markers of Adenomatous Polyps (MAP) study I and II

The Markers of Adenomatous Polyps I and II (MAP I and II) cross-sectional studies were conducted by the same principal investigator (RMB) using almost identical protocols. The methods for the two studies are described in detail elsewhere (79-81). Participants in the MAP I study (n=474) were recruited from community gastroenterology practices in Winston-Salem and Charlotte, North Carolina, while those in the MAP II study (n=233) were recruited among patients who received care at Consultants in Gastroenterology, PA, a large, private practice clinic in Columbia, South Carolina. Eligible subjects were persons 30-74 years of age with no prior

history of colorectal neoplasms who were scheduled to undergo elective, outpatient colonoscopy. The study eligibility criteria in the two studies were identical. Cases (n = 233) were subjects diagnosed with an incident colon or rectal adenoma at the time of their colonoscopy procedure. Controls (n = 312) were all subjects free from all types of polyps during colonoscopy. Subjects with hyperplastic polyps (n = 70) were defined as missing information on the outcome of interest.

Data Collection

Questionnaire data

Before undergoing the colonoscopy procedure, all consenting subjects completed a mailed questionnaire regarding their demographic characteristics, medical history, use of medications, and habits. This questionnaire included, but was not limited to, smoking and drinking history, aspirin and non-steroidal anti-inflammatory (NSAID) use, and physical activity (PA). In addition, diet and nutritional supplement intakes were assessed through a modified 153-item Willet Food Frequency Questionnaire. Intakes of polyunsaturated fatty acids (PUFA), vitamin C, and fiber, were estimated as a total of diet and supplement use (144, 145).

Biomarker data

All participants provided blood samples that were drawn into red-coated, pre-chilled Vacutainer tubes, plunged into ice, and protected from light sources. Upon immediate delivery to the lab, the samples were centrifuged under refrigeration, after which plasma and serum were separated and aliquotted into O-ring-capped, amber-colored cryopreservation vials filled with inert gas. The samples were stored at -70°C until analysis.

All biomarker analyses were performed by the Molecular Epidemiology and Biomarker Research Laboratory at the University of Minnesota (Minneapolis, MN). Plasma FIP concentrations were measured using a gas chromatography-mass spectrometry (GC-MS) method

(114). The FIP were extracted from the participants' samples using deuterium (4)-labeled 8-iso-prostaglandin F_2 alpha as an internal standard with unlabeled, purified F_2-isoprostane as a calibration standard. FOP were measured in plasma samples by the modified Shimasaki method as previously described (39, 115). A mixed solution of plasma and ethanol/ether was centrifuged for 10 minutes at 3,000 rpm after which 1.0 ml of supernatant was added to cuvettes for spectrofluorometric readings. Relative fluorescence intensity in units per milliliter of plasma at 360/430nm wavelengths was calculated by a spectrofluorometer. Quinine sulfate diluted in 0.1 N H_2SO_4 was used for calibration. Approximately 22% of the samples were serum rather than plasma due to limited supply of plasma in the study population. This was not expected this to affect the FOP measurement or analysis as there was a high level of correlation between a random sample of subjects on whom we measured FOP in both serum and plasma ($r = 0.9$; $p < 0.001$). High sensitive CRP was measured by latex-enhanced immunonephelometry on a Behring nephelometer II (BN-II) analyzer (inter-assay CV 4%; Behring Diagnostics, San Jose, CA). MtDNA was analyzed in 2013 by real-time quantitative polymerase chain reaction (qPCR) as described previously (55). For the qPCR method we used two primers, one for MtDNA and one for nuclear DNA. The ratio of MtDNA to nuclear DNA was determined by serial dilution of a healthy referent genomic DNA sample.

Structural Equation Model and Statistical Methods

All SEMs were analyzed with the Mplus software version 7.1 statistical package, which was developed specifically for these types of analyses (83). Other analyses were performed with SAS software version 9.3 (146). All missing data on outcome variables were assumed to be missing at random (Mplus default setting) (147). Due to the racial homogeneity of the study populations, all analyses were restricted to non-Hispanic whites. A two-sided p-value <0.05 indicated statistical significance.

All SEM analyses originated from the baseline model in which we theorized one continuous latent factor (here onward termed oxidative stress) composed of five continuous indicator variables chosen *a priori*: FIP, FOP, CRP, MtDNA, and Gtoc. The baseline model included a direct effect from oxidative stress to colorectal adenoma, as well as the following potential confounding variables: age, sex, BMI, smoking, drinking, aspirin use, non-steroidal anti-inflammatory (NSAID) use, and a variable designating study population (Figure 3.1). Oxidative stress was standardized (mean=0, variance=1) in order to estimate loading factors for all five biomarkers. To aid empirical identification, the continuous indicator variables for oxidative stress were mathematically transformed to have similar variances; this transformation did not affect standardized results. Unless stated otherwise, all models were estimated using a generalized linear *logit* SEM under maximum likelihood estimation. A drawback of this method is that it does not provide usual SEM fit statistics (root mean squared error of approximation (RMSEA), the comparative fit index (CFI), the Tucker-Lewis index (TLI), and the standardized root mean square residual (SRMR)).

Before pooling the data on the two study populations, we first tested whether the construction of the latent oxidative stress variable in the baseline model differed between the MAP I and MAP II datasets, and between men and women. This approach is called invariance testing and it is done by setting unstandardized loading factors (λ_u) equal across groups (117, 118, 143). Generalized linear *probit* models, under weighted least squares estimation (WLSMV), were used for testing invariance because they provide SEM fit statistics in addition to nested χ^2 difference tests for binary outcomes. The results of these tests influenced the decision to model direct effects of sex and study population on each biomarker rather than on oxidative stress (Figure 3.1).

The following three models were used to estimate the association between oxidative stress and colorectal adenoma. Each subsequent model added more covariates. All ordinal variables were categorized by sex-specific tertile cut points.

Model 1: The baseline model of oxidative stress and colorectal adenoma included the following covariates: age (continuous, years), sex (binary), BMI (<25, 25-29.9, ≥30 kg/m^2), smoking (never, former, current), drinking (never, former, current), aspirin use (binary), non-steroidal anti-inflammatory (NSAID) use (binary), and study population (binary); n = 526.

Model 2: The partially adjusted model included the variables from the baseline model as well as the following: reported physical activity (ordinal), total energy intake (ordinal), total fiber intake (ordinal), plasma cholesterol levels (continuous, mg/dl); n = 469.

Model 3: The fully adjusted model included total vitamin C intake (ordinal), and total poly-unsaturated fatty acid intake (PUFA) (ordinal) in addition to the covariates in the partially adjusted model; n = 469.

In addition to the estimated direct effect of oxidative stress on colorectal adenoma, we assessed model-specific direct effects of exposure on oxidative stress. We also estimated and reported indirect effects for exposures on colorectal adenoma, mediated by oxidative stress (148). The estimated standardized direct effect from oxidative stress to colorectal adenoma is interpreted as the change in odds of colorectal adenoma, per one standard deviation (s.d.) change in oxidative stress. Estimated direct effects from exposures to oxidative stress (γ_s) are interpreted as the s.d. change in oxidative stress per one unit change in exposure. Indirect effect estimates are interpreted as the change in odds of colorectal adenoma per one unit change in exposure, mediated by oxidative stress.

On the basis of prior evidence and biologic theory, we considered an alternative model with two oxidative stress latent variables representing two different processes, possibly oxidative stress and inflammation (oxstress1, oxstress2) (Figure 3.2). To identify the two latent variables, FIP was restricted to load to only one of the latent constructs. We repeated the model changing the restriction from FIP to Gtoc, and then again from Gtoc to CRP to see if the results changed. The model with two-latent factors was examined to assess whether each factor had a distinct biomarker pattern with little or no cross-loadings.

RESULTS

Selected baseline characteristics of the MAP I and II study populations are presented in Table 3.1. The average age of the combined study population was 56.8 years. Fifty-three percent of the subjects were women, and the average BMI was 28.0 kg/m^2. Most of the participants were never or former smokers (77%) and half were current drinkers (50.8%).

The tests for invariance revealed that the construction of the latent oxidative stress variable differed by sex and by study. Among women, the unstandardized loading factors for FIP and CRP were higher and statistically significantly different from the respective loading factors for men (Table 3.2). This difference however, did not result in a meaningful change in the characterization of oxidative stress on the standardized scale as estimates (λ_s) for FIP, Gtoc, and CRP were fairly similar between men and women (0.51, 0.69, 0.27 respectively for men; and 0.61, 0.68, 0.49 respectively for women). The test for study invariance revealed that in addition to FIP and CRP, the loading factor for MtDNA varied between the two study populations: the MtDNA standardized loading was -0.27 in the MAP II study population, compared to 0.05 in the MAPI population. Instead of using stratification, we allowed the variables for sex and study population to have direct effects on each biomarker rather than on oxidative stress (Figure 3.1).

The baseline model, under a *probit* distribution, demonstrated good fit (χ^2 p-value = 0.91; RMSEA <0.01; CFI = 1.00; TLI = 1.06).

In all three models, oxidative stress was strongly, and positively associated with colorectal adenoma (Table 3.3). In the baseline model, there was a twofold change in the odds of colorectal adenoma for each s.d. increase in oxidative stress with an estimated odds ratio (OR) of 1.96 and a 95% confidence interval (CI) from 1.19 to 3.24. This OR point estimate was 2.56 (95% CI: 1.27-5.17) in the partially adjusted model, and increased further to 2.61 (95% CI: 1.25-5.46) in the fully adjusted model.

Latent oxidative stress was characterized by positive, statistically significant standardized loading factors from FIP (λ_s = 0.61, 95% CI: 0.51-0.71), Gtoc (λ_s = 0.57, 95% CI: 0.47-0.67), and CRP (λ_s = 0.40, 95% CI: 0.30-0.50); MtDNA and FOP loading factors were null (Table 3.4). This result did not change meaningfully in either the partially or fully adjusted models.

Across all three models, several exposure variables were associated with oxidative stress. From the adjusted models, higher amounts of physical activity, total fiber, and vitamin C intakes were estimated to have negative effects on oxidative stress; the latter two were statistically significant (Table 3.4). Conversely, higher cholesterol, PUFA, and total energy intakes were estimated to have positive effects on oxidative stress; all were statistically significant. We used the baseline model to estimate model-defined indirect effects from exposure to colorectal adenoma, mediated by oxidative stress. BMI was estimated to have a positive, statistically significant indirect effect on colorectal adenoma (for BMI ≥30: OR=2.64, 95% CI: 1.22-5.69; for BMI 25-29.9: OR=1.44, 95% CI: 1.03-2.01) (Table 3.3). Estimates for aspirin use, and alcohol consumption were inverse but more modest, and estimates for smoking were null.

We explored three different two-factor models to assess whether the five biomarkers under study could identify two different oxidative stress constructs (oxstress1, oxstress2). After

allowing FIP to only load to oxstress1, CRP had a statistically significant cross-loading with both factors (λ_s = 0.25, p-value = 0.03 for oxstress1; λ_s = 0.66, p-value = 0.04 for oxstress2). When CRP was restricted to only oxstress1, Gtoc had a meaningful and but not statistically significant cross-loading (λ_s = 0.26, p-value = 0.07 for oxstress1; λ_s = 0.45, p-value < 0.01 for oxstress2). When Gtoc was restricted to oxstress1, no loading factors were significant for oxstress2.

DISCUSSION

In the two pooled data from two cross-sectional studies, a latent oxidative stress variable had a strong and statistically significant positive association with newly diagnosed colorectal adenoma. Oxidative stress was characterized by positive loading factors from F_2-isoprostanes, γ-tocopherol, and C-reactive protein. Additionally, a number of pro-/antioxidant exposure variables were statistically significantly associated with oxidative stress in the theorized directions, and were indirectly associated with colorectal adenoma through oxidative stress. To our knowledge, this is the first study to use SEM with latent variables to investigate the association between oxidative stress and a health outcome.

Numerous biomarkers of oxidative stress have been proposed and used in research over the last few decades (6, 36). Each of these biomarkers usually reflects a specific aspect of *in vivo* oxidation without offering an overall measure of oxidative stress (111, 112). A latent variable can overcome this problem by reducing the measurement error related to any single biomarker by modeling the shared covariance across multiple measures (26, 27). SEMs allow constructing systems of structural equations where observed variables can either cause, or be causes of, the latent variable; however, they do not obviate the need for sound biologic theory.

With the use of SEM we identified a single latent variable, termed oxidative stress, primarily constructed from three (FIP, CRP, and Gtoc) of five (also FOP and MtDNA) *a priori* selected plasma biomarkers. The characterization of oxidative stress was similar to our previous

analysis, in a slightly different dataset, that did not include the health outcome of colorectal adenoma (143). By accurately reflecting lipid peroxidation, FIP serves as an established biomarker of oxidative stress (37, 40, 120). Basic science evidence supports a close link between oxidative stress and inflammation (65), and while CRP is recognized as a non-specific biomarker of inflammation, it is correlated with FIP and other biomarkers of oxidative stress (66-69). Circulating levels of Gtoc are thought to be more indicative of underlying oxidative stress conditions rather than dietary intake (57, 60), and evidence supports a positive association between Gtoc and oxidative stress biomarkers (43, 62). FOP and MtDNA, recently proposed as markers of oxidative stress (38, 47, 48, 56), did not co-vary with FIP, CRP, or Gtoc in a meaningful way, and therefore did not contribute to the characterization of the latent variable in our study population.

In this study, the association between the latent oxidative stress variable and colorectal adenoma was stronger than what is currently reported for the individual biomarkers. There are a few inconsistent reports of the association between FIP and colorectal adenoma. Using binary FIP from sex-specific median cut points, Kong et al. found the plasma biomarker to be positively associated with adenoma reporting an odds ratio of 1.89 (18). In contrast, Siamakpour-Reihani et al. reported null associations between four different urinary isoprostanes with colorectal adenoma in a prospective cohort study (41). This discrepancy may be in part due to the differences in study design. The Siamakpour-Reihani study measured FIP approximately 10-15 years before colonoscopy, while Kong et al. measured FIP shortly before colonoscopy. Additionally, the disagreement between the two studies may be attributed to the differences in the analytic medium - spot urine samples in the Siamakpour-Reihani study versus plasma by Kong. According to Halliwell and Lee, spot plasma and urine measurements of FIP are not interchangeable, and plasma samples are generally preferred (120). Similar to FIP, data on the association between plasma Gtoc and colorectal adenoma are sparse. One study reported higher levels of Gtoc in

persons with adenoma relative to sigmoidoscopy-conformed controls ; however that result was no longer statistically significant when fully adjusted for confounding variables (63). Unlike Gtoc and FIP, CRP has been more frequently studied in relation to colorectal neoplasia. A meta-analysis of prospective studies that investigated the association of CRP with colorectal cancer found a weak but statistically significant odds ratio of 1.12. There are fewer studies of CRP with adenoma and those studies have reported mostly null results (18, 76-78).

While our study results differ from those reported elsewhere, the discrepancy has a plausible explanation. It is possible that each biomarker is weakly associated with adenoma via a common cause. If a latent variable measures that common cause (in this case oxidative stress) through the shared covariance of the biomarkers, a stronger association is expected.

In addition to our primary finding, our baseline model estimated that some variables were indirectly associated with adenoma, mediated by oxidative stress. In our study, the inverse associations with colorectal adenoma through oxidative stress for both regular aspirin use, and alcohol consumption, were borderline statistically significant. Aspirin use is a preventive agent against adenoma (149), and is associated with lower oxidative stress biomarkers (108), which is consistent with our results. The reports on the association between alcohol consumption and oxidative stress are mixed (43, 128), but heavy alcohol use was found to be associated with higher risk of adenoma (85). This evidence does not support our result and could be due to interactions or correlations with smoking, and or diet, that we did not model. Among all exposure factors in the present study, higher BMI was most strongly associated with colorectal adenoma, mediated by oxidative stress. This observation is in agreement with the current literature, which identifies obesity as an established risk factor for adenoma (150, 151) and demonstrates a link between obesity and oxidative stress through specific metabolic pathways (107). Our results further raise the possibility that the positive association between obesity and colorectal adenoma is mediated by an oxidative stress pathway.

While oxidative stress is a complex multifaceted process we were unable to model it in more than one latent construct. With a two-factor SEM, there were consistent biomarker cross-loadings, indicating that the two constructs were not adequately distinguishable. This result does not exclude the possibility of multiple oxidative stress processes working in a complex network, but it does indicate that the five biomarkers in the present study were unable to identify more than one latent variable.

Perhaps the most important limitation of the current study is the use of prevalent, albeit newly diagnosed, adenoma cases coupled with required SEM assumptions. This design feature precludes clear understanding of the sequence of events, but temporality of variables must be assumed in SEM. Because collection of exposure information happened before the patient underwent a colonoscopy, knowledge of adenoma status did not affect the subjects' questionnaire responses. However, changes in blood measurements could have occurred after adenoma development. The SEM method must assume the temporal and causal relation among variables, and there is no test or assurances that this assumption is met. To obtain a sufficient sample size, we pooled data across two different study populations and did not stratify on gender. The results of the invariance tests indicated that latent oxidative stress was only partially invariant between the two studies, and between men and women. It is possible our results could differ if stratified with larger sample sizes. Moreover, as almost all participants in the MAP I and II studies were non-Hispanic whites, we cannot generalize the results to other racial or ethnic groups. Finally, we cannot be certain that our latent variable is exclusively, or wholly, a construct of oxidative stress. This is a limitation of SEM that can be addressed only through models built with sound biologic theory, which we did to the best of our ability.

Future studies in this field should consider the noted limitations of the present study. As colorectal neoplasms tend to grow slowly, only a prospective design, preferably with multi-year follow-up, is suitable for determining the temporal relation between blood biomarkers and tumor

initiation. Also, given our findings from gender invariance testing, future research should consider the variability of oxidative stress in men and women, and account for that in sample size estimates. Studies with more biomarkers of oxidative stress would allow for better characterization of the construct and its association with colorectal adenoma. Such new studies might identify multiple oxidative stress constructs, which was not possible with the current data. Lastly, the logical next step would be to investigate the association between a latent oxidative stress variable and colorectal cancer.

In conclusion, this novel application of SEM has provided further evidence that colorectal adenoma is positively associated with oxidative stress. This method has the ability to reduce individual variability of a single biomarker by modeling the shared covariance across multiple measures. Further use of SEM analysis in colorectal adenoma and carcinoma studies could assist in understanding the role of oxidative stress in disease progression. Based on the experience in the current study, the SEM methodology can be applied to studying other oxidative stress-related health outcomes. It can also be extended to other areas of research where it is necessary to combine different, but imperfect measurements to describe a complex biologic phenomenon.

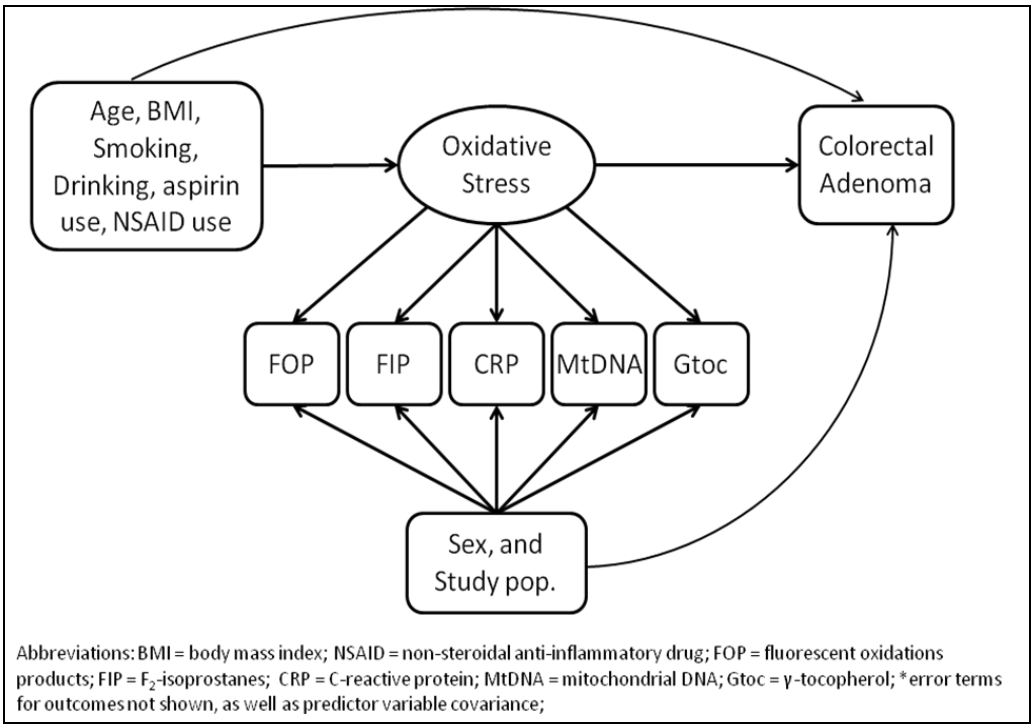

Figure 3.1. Graphical representation of baseline structural equation model

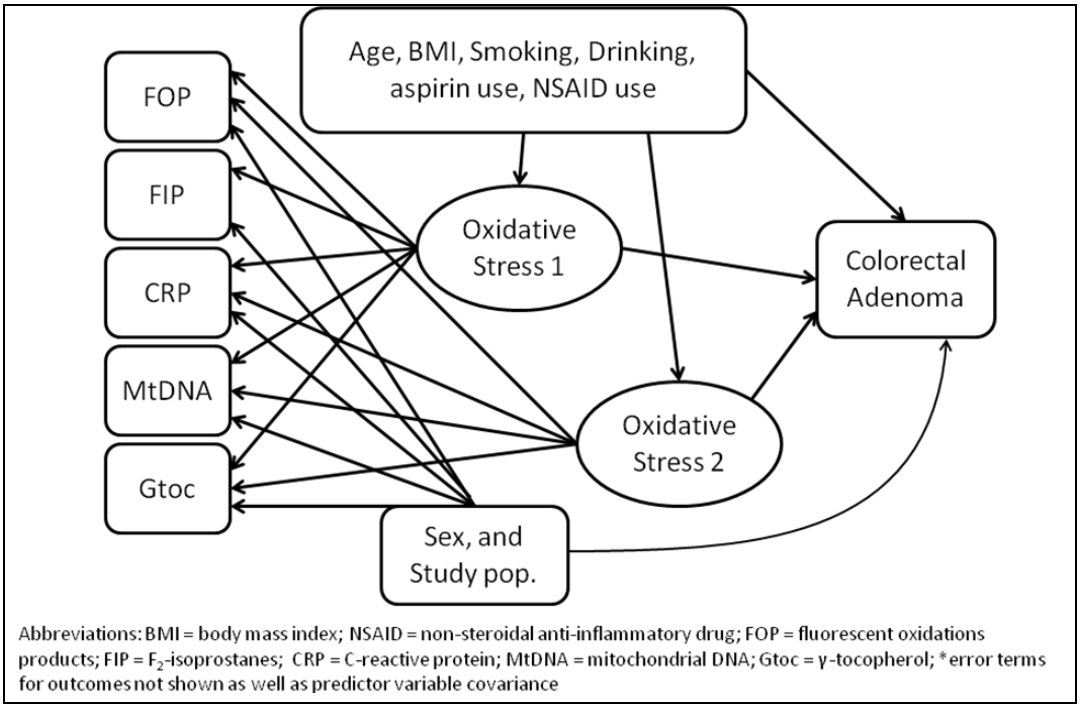

Figure 3.2. Graphical representation of alternative structural equation model with 2 latent oxidative stress variables

Table 3.1. Selected baseline characteristics of non-Hispanic white subjects in the Markers of Adenomatous Polyps I and II datasets

	Cases (n=188)		**Controls (n=255)**		**Missing (n=83)**		**Total (n=526)**	
	μ or n	% or s.d.	μ or n	% or s.d.	μ or n	% or s.d.	μ or n	% or s.d.
Age, years	58.0	8.1	55.6	9.4	57.8	8.7	56.8	8.9
Sex (% women)	77	41.0	163	63.9	35	42.2	275	52.3
BMI, kg/m^2	28.0	6.4	27.6	6.1	29.0	6.6	28.0	6.3
Phys. act., MET-hours/week	187	138	189	131	209	125	191	133
Smoker								
Never	50	26.6	128	50.2	22	26.5	200	38.0
Former	78	41.5	95	37.3	32	38.6	205	39.0
Current	60	31.9	32	12.6	29	34.9	121	23.0
Drinker								
Never	43	22.9	87	34.1	23	27.7	153	29.1
Former	43	22.9	46	18.0	17	20.5	106	20.2
Current	102	54.3	122	47.8	43	51.8	267	50.8
Aspirin use	68	36.2	88	34.5	32	38.6	188	35.7
NSAID use	41	21.8	90	35.3	22	26.5	153	29.1
Dietary Intake								
Total energy, kcal	2000	798	1830	892	2051	921	1925	868
PUFA, gm	14.2	6.8	13.6	9.1	14.5	7.2	13.9	8.1
Vitamin C, mg	269	348	269	309	254	282	266.4	319
Fiber, gm	21.5	9.6	20.2	11.0	22.3	14.1	21.0	11.1
Plasma levels								
Cholesterol, mg/dl	203.6	35.6	204.2	39.8	208.4	39.7	204.7	38.3
FIP, pg/ml	90.2	48.9	84.6	35.5	85.2	34.9	86.8	40.9
FOP, avg. std.ref.adj.	0.05	0.11	0.04	0.02	0.06	0.14	0.05	0.09
CRP, μg/ml	5.3	6.0	4.7	5.9	5.8	6.3	5.1	6.0
MtDNA,rel.nucl.DNA	2.2	2.9	2.8	2.3	2.0	1.8	2.5	2.5
Gtoc, mg/dl	0.23	0.11	0.20	0.11	0.21	0.10	0.21	0.11

Symbols and abbreviations: μ = mean; s.d. = standard deviation; MET = metabolic equivalent; NSAID = non-steroidal anti-inflammatory drug; PUFA = poly-unsaturated fatty acid; FIP = F_2-isoprostanes; FOP = fluorescent oxidation products in average standard reference adjusted; CRP = C-reactive protein; MtDNA = mitochondrial DNA in relative MtDNA to nuclear DNA; Gtoc = γ-tocopherol;

Table 3.2. Test for oxidative stress weak factorial invariance by sex MAP I and II study populations

No Invariance

	Among Men (n = 251)				Among Women (n = 276)					
	λ_u	95% LCI	95% UCI	λ_s	λ_u	95% LCI	95% UCI	λ_s	Model Fit Statistics	
MtDNA	-1.03	-1.87	-0.19	-0.20	-1.28	-2.13	-0.42	-0.27	$\chi^{2\ pvalue}$	0.26
FOP	0.07	-0.20	0.33	0.06	0.19	-0.68	1.05	0.05	RMSEA	0.02
FIP	1.46	0.88	2.03	0.50	3.15	2.45	3.85	0.61	CFI	0.96
CRP	1.11	0.43	1.79	0.26	2.38	1.63	3.13	0.49	TLI	0.94
Gtoc	3.70	2.36	5.05	0.73	3.39	2.60	4.19	0.67		

Partial Invariance

	Among Men (n = 251)				Among Women (n = 276)					
	λ_u	95% LCI	95% UCI	λ_s	λ_u	95% LCI	95% UCI	λ_s		
MtDNA	-1.18	-1.79	-0.57	-0.23	-1.18	-1.79	-0.57	-0.25	$\chi^{2\ pvalue}$	0.33
FOP	0.08	-0.18	0.34	0.07	0.08	-0.18	0.34	0.02	RMSEA	0.02
FIP	1.47	0.91	2.04	0.51	3.14	2.43	3.85	0.61	CFI	0.98
CRP	1.15	0.45	1.85	0.27	2.38	1.63	3.13	0.49	TLI	0.96
Gtoc	3.45	2.77	4.12	0.69	3.45	2.77	4.12	0.68		

Symbols and abbreviations: MAP = Markers of Adenomatous Polyps; λ_u = unstandardized loading factor; λ_s = standardized loading factor; LCI = lower confidence interval; UCI = upper confidence interval; χ^2 = Chi-squared; RMSEA = root mean squared error of approximation; CFI = comparative fit index; TLI = Tucker-Lewis index; FIP = F_2-isoprostanes; FOP = fluorescent oxidation products; CRP = C-reactive protein; MtDNA = mitochondrial DNA; Gtoc = γ-tocopherol;

Table 3.3. Model estimated direct and indirect effects on colorectal adenoma

	OR_s	95% LCI	95% UCI	p-value
Oxidative stress direct effects				
Baseline model[a]	1.96	1.19	3.24	0.01
Partially adjusted model[b]	2.56	1.27	5.17	0.01
Fully adjusted model[c]	2.61	1.25	5.46	0.01
Indirect effects mediated by oxidative stress[d]				
BMI (kg/m^2)				
<25	referent			
25-29.9	1.44	1.03	2.01	0.03
≥30	2.64	1.22	5.69	0.01
Smoking				
Never	referent			
Former	1.02	0.84	1.22	0.86
Current	1.23	0.95	1.59	0.13
Alcohol Drinking				
Never	referent			
Former	0.76	0.56	1.04	0.09
Current	0.73	0.54	0.99	0.05
Aspirin use	0.79	0.62	1.01	0.06

Symbols and abbreviations: OR_s= standardized odds ratio; LCI = lower confidence interval; UCI = upper confidence interval; BMI = body mass index; [a]adjusted for age, sex, BMI, smoking, alcohol consumption, aspirin, and non-steroidal anti-inflammatory use; [b]additionally adjusted for physical activity, total fiber intake, total energy intake, and plasma cholesterol; [c]additionally adjusted for total vitamin C intake, and poly-unstaturated fatty acid intake; [d]estimated from baseline model and not standardized

Table 3.4. Model estimated effects on latent oxidative stress

	Baseline model			
	γ_s	95% LCI	95% UCI	p-value
Age[a]	-0.004	-0.017	0.009	0.58
BMI (kg/m^2)				
<25	referent			
25-29.9	0.54	0.28	0.80	<0.01
≥30	1.44	1.18	1.69	<0.01
Smoking				
Never	referent			
Former	0.02	-0.25	0.30	0.86
Current	0.30	0.00	0.61	0.05
Alcohol Drinking				
Never	referent			
Former	-0.41	-0.75	-0.06	0.02
Current	-0.46	-0.75	-0.18	<0.01
Aspirin use	-0.34	-0.58	-0.11	0.01
NSAID use	-0.06	-0.31	0.19	0.62
	Partially adjusted model			
Physical activity[b]	-0.11	-0.24	0.01	0.08
Fiber[b]	-0.42	-0.58	-0.25	<0.01
Total energy[b]	0.21	0.05	0.37	0.01
Cholesterol[a]	0.009	0.006	0.011	<0.01
	Fully adjusted model			
Vitamin C[b]	-0.34	-0.48	-0.21	<0.01
PUFA[b]	0.17	0.02	0.33	0.03
	Loading factors for oxidative stress[c]			
	λ_s	95% LCI	95% UCI	p-value
MtDNA	-0.05	-0.15	0.05	0.34
FIP	0.61	0.51	0.71	<0.01
FOP	0.01	-0.11	0.13	0.84
CRP	0.40	0.30	0.50	<0.01
Gtoc	0.57	0.47	0.67	<0.01

Symbols and abbreviations: MAP = Markers of Adenomatous Polyps; γ_s = standardized effect estimate; λ_s = standardized loading factor; LCI = lower confidence interval; UCI = upper confidence interval; BMI = body mass index; PUFA = poly-unsaturated fatty acid; FIP = F_2-isoprostanes; FOP = fluorescent oxidation products; CRP = C-reactive protein; MtDNA = mitochondrial DNA; Gtoc = γ-tocopherol; [a]continuous predictor; [b]tertile predictor; [c]estimates from the baseline model;

CHAPTER 4. ASSESSING THE CAUSAL ASSUMPTIONS AND METHODOLOGICAL CHALLENGES OF STRUCTURAL EQUATION MODELS IN EPIDEMIOLOGIC STUDIES OF BIOLOGIC CONSTRUCTS

ABSTRACT

Structural equation models (SEMs) can be used to provide unique research perspectives for investigating biologic phenomena measured by imperfect biomarkers. SEMs can be used to model biologic constructs through the shared covariance of multiple biomarkers in simplified causal models. The drawback of SEM is its dependence on multiple assumptions, some of which may be quite strong. In this communication, we explicate some of the model assumptions that may be common in SEM-based studies of biologic phenomena. We also investigate potential error from information, selection, and confounding biases. In SEMs, these biases can manifest themselves in ways that differ from those when using traditional regression analyses, and therefore must be appreciated differently. For each bias we simulate data according to a specified model, and analyze the direction and magnitude of error. It is important to critically evaluate the assumptions and biases of any SEM. For studies of biologic phenomena, the SEM assumptions will likely center around the latent variable and the measured biomarkers that identify it. Studies that employ a cross-sectional or case-control study design require careful attention to the possibility of reverse causation. Sound biologic theory can best guide SEMs used to investigate biologic phenomena.

INTRODUCTION

Epidemiologic studies often assess the effects of a biologic phenomenon on a health outcome, or conversely the effects of an exposure on the biologic phenomenon. Examples of the former are studies of the effects of inflammation on cardiovascular disease (152); examples of the latter are studies of the possible effects of body mass index on oxidative stress (38, 43). The biologic phenomena of interest in these studies are usually examined by measuring the corresponding biomarkers (e.g., cytokines for inflammation or products of oxidation for oxidative stress). While biomarkers are commonly used in health research, often they are only imperfect measurements of the biologic phenomenon of interest. Structural equation models (SEMs) with latent variables offer a possible way to study such imperfectly measured phenomena by using multiple biomarkers to characterize the phenomenon indirectly (26, 27).

Couched in the counterfactual definition of causality, SEMs are functional causal models that allow for relatively easy statistical computation of theorized causal pathways (26, 35). SEMs are composed of a system of structural equations that determine how a dependent variable is affected by its direct parents. When multiple equations link together, the structural equations are called a structural model (35). The structural model incorporates latent variables (unobserved, or poorly measured) by identifying them through their theorized causal relationships with observed variables (26). The resulting SEM and its parameters can be estimated if parametric underlying distributions are assumed for each dependent variable. These assumptions then allow for the simultaneous maximization of the combined likelihoods and estimation of model parameters.

The estimated parameters in SEMs have causal interpretations within the model (26, 35) which often take the form of a “direct effect” or “indirect effect” of an explanatory variable on an outcome variable (26). Substantial published research addresses the identification and estimation of direct and indirect effects, including the required assumptions for valid estimation (29-31, 34,

148, 153, 154). However, the degree to which the SEM assumptions are met (155), and the effect of violations of these assumptions has not been extensively addressed for epidemiologic applications.

In this communication, we use Monte Carlo simulations to explicate model assumptions and methodological challenges in SEMs when applied to epidemiologic research concerning biologic phenomena. Specifically, we summarize the assumptions commonly made when SEMs are used to identify a latent variable that represents a particular biologic construct. We also consider potential biases and their impact on the results of SEM-based studies. For each of these issues, we provide a short background, describe the model used to quantify the strength of the assumption and the magnitude of bias, present analyses and results, and offer conclusions. We also provide recommendations for future research. Throughout the communication, we use a specific example of a previously completed cross-sectional study that examined oxidative stress as the biological phenomenon of interest (see chapters 2 and 3). All Monte Carlo simulations were conducted using Mplus version 7.1 (83).

SEM ASSUMPTIONS

SEMs with latent variables allow evaluating biologic phenomena that are not directly observed (i.e., constructs) by modeling the shared covariance of multiple measures (e.g., biomarkers); however, the analytical power of SEMs comes at the cost of having to make more assumptions than are required for traditional regression analyses. SEMs are closely related to causal diagrams (also known as directed acyclic graphs - DAGs) and the two approaches share similarities in their causal assumptions (35, 155). As in DAGs, each arrow in a SEM represents an explicit causal assumption about the two variables that begin and end that arrow (26, 35, 82). Further, every arrow omitted from a SEM or causal diagram also represents an explicit causal

assumption. In this section, we explore some specific SEM assumptions that might be made when modeling a biologic construct.

SEM-based studies of biologic constructs typically involve assumptions that the latent variable is causally related to its determinants (*e.g.,* antioxidant exposures) and its biomarkers, and also causally affects the health outcome. For example, Figure 4.1a represents a very basic SEM involving one independent variable (body mass index – BMI), one latent variable (oxidative stress), the primary health outcome variable (colorectal adenoma) and five indicator variables that are biomarkers of oxidative stress (F_2-isoprostanes – FIP, fluorescent oxidation products – FOP, C-reactive protein – CRP, γ-tocopherol – Gtoc, and mitochondrial DNA copy number – MtDNA copies). For the purposes of this example, assume that each variable in the SEM precedes its effect, and that there is no measurement error, selection bias, or unmeasured confounding influencing the causal effects of interest: BMI on oxidative stress, and oxidative stress on colorectal adenoma.

Even the relatively simple SEM (shown in Figure 4.1a) has many additional assumptions, such as absence of additional effects (depicted in Figure 4.1b), the appropriateness and plausibility of which need evaluation. For example, the diagram in Figure 4.1a assumes the absence of any direct effect of BMI on colorectal adenoma (Figure 4.1b). Without the direct effect, BMI can only affect adenoma through the mechanism involving oxidative stress, and any observed correlation in the study data between BMI and adenoma must come from that pathway. Another assumption reflected in Figure 4.1a is that BMI only has indirect effects on the biomarkers of oxidative stress. However, as shown in Figure 4.1b it may be plausible that BMI has an effect on CRP not through the mechanism of oxidative stress. Also, in Figure 4.1a, each biomarker is indirectly associated with colorectal adenoma and with other biomarkers because oxidative stress is a common cause. However, it might be possible for a biomarker to associate with adenoma, or for two biomarkers to be related to each other by a mechanism other than

oxidative stress (Figure 4.1b). In Figure 4.1a, any correlation between the biomarkers, and between the biomarkers and adenoma, is assumed to be from oxidative stress. Thus, the assumptions reflected in Figure 4.1a may be appropriate, but their plausibility must be assessed, and compared with other possibilities.

Another important assumption concerns the relationships between the latent oxidative stress variable and each biomarker. In the SEM context, both Figures 4.1a and 4.1b reflect the assumption that the shared covariance of the five biomarkers (FIP, FOP, CRP, MtDNA, and Gtoc) can adequately characterize oxidative stress. Essentially, the measurement of the latent oxidative stress variable is limited by information in these five biomarkers. As a large number of oxidative stress biomarkers are used in research (6, 112, 156), it is possible that additional biomarkers could contribute further information and provide a better measurement of oxidative stress. For these reasons each latent variable in a SEM requires justification *via* critical evaluation of its components.

In the next sections of this communication, we evaluate some of the model assumptions that may be common in SEM-based studies of biologic phenomena such as oxidative stress. We also investigate potential errors or problems with interpretation of findings resulting from failure to determine the sequence of events (temporality), information bias (misclassification or measurement error), selection bias, and uncontrolled confounding.

TEMPORALITY AND INFORMATION BIAS

A necessary requirement for the assumption of causality is that each cause must precede its effect. In many observational, especially cross-sectional, studies this requirement is not assured. This is a particular concern in cross-sectional studies that involve biomarker measurements and health outcomes. For instance, in the previously mentioned (see chapter 3) cross-sectional study of plasma biomarkers of oxidative stress and their relation to colorectal

adenoma, all variables were measured at approximately the same time. In particular, the biomarkers reflected oxidative activity at the time of, or just before their measurement, whereas adenomas developed over a longer period of time. It is therefore likely that the outcome occurred before the biomarker (exposure) was measured. If the cross-sectional observations are to be used to estimate an effect of oxidative stress on adenoma, the measured oxidative stress biomarkers must serve as a viable proxy for corresponding biomarkers that preceded the development of adenoma. An imperfect correlation of the measured biomarkers with earlier biomarker levels is a possible source of measurement error.

Another issue that can arise in cross-sectional studies is the possibility that the health outcome, in this case adenoma, could cause an increase in oxidative stress; this would represent a violation of the causal assumptions used in the SEM, and is referred to as "reverse causation". If oxidative stress during an earlier period is a cause of adenoma and then the adenoma later affects the subsequent level of oxidative stress, the measurement error will be differential. Reverse causation is, where oxidative stress_1 precedes and causes adenoma, and adenoma then affects subsequent, measured oxidative stress_2. The degree of measurement error will differ depending on the presence or absence of adenoma.

Exposures may have also changed between the time adenoma developed and the time of measurement. For instance, changes in weight may occur after the development but prior to the diagnosis of colorectal tumors (although this is less likely for adenoma). Similar to the situation for biomarkers, information on exposures must be correlated with earlier exposures (ρ Ox stress, ρ BMI) that were present during a causally relevant time to be a viable proxy variable. If a variable is categorical, measurement error is called misclassification.

Measurement error and misclassification are important because they can cause information bias—erroneous effect estimates that result from errors in measurement (157). To

quantify the potential magnitude of information bias, we assume different degrees of measurement error due, at least in part, to having measured exposure, oxidative stress, and colorectal adenoma at the same time, without accounting for temporal changes. We quantify potential information bias due to measurement error, misclassification, and reverse causation. We consider an example involving causal effects of body mass index (BMI) (exposure), oxidative stress (intermediate), and colorectal adenoma (outcome). Using simulation analyses we first consider the true causal model, and then the models that reflect different scenarios involving measurement error.

True causal model

Figure 4.2 represents the assumed causal relationships between colorectal adenoma (binary), oxidative stress (Oxstress, continuous), and BMI (binary). The subscripts (1, 2, 3) indicate time at measurement. Time 3 represents when all three variables were measured (as in a cross-sectional study), while time 2 and time 1 are points in the past that were causally relevant for oxidative stress and adenoma. In this example, we assume the DAG represents the true, causal relationships, and that there is no bias due to non-random selection or unmeasured confounding. We also assume there is no additional measurement error other than that we explicitly represented. For simulation purposes, we treat oxidative stress as a measured variable rather than a latent construct. If oxidative stress were not measured directly, but rather only assessed based on biomarkers, we would expect the same pattern of biases.

Information bias

The model in Figure 4.2 is used to simulate measurement error when reverse causation is not an issue. In this simulation, the population effects of interest were those for BMI_1 on $Oxstress_2$, BMI_2 on $Oxstress_3$, and $Oxstress_2$ on $Adenoma_3$. The population effects were set to the values depicted in the figure (1.25 standard deviation change [s.d.] in oxidative stress per unit of

BMI, 1.25 s.d. change in oxidative stress per unit of BMI, and a 2.0 increase in odds of adenoma per s.d. change in oxidative stress). The causal effects of BMI_1 on BMI_2 and BMI_2 on BMI_3 represented weight gain or loss during the time period in which adenoma developed, and as a result, potential misclassification of BMI_3; stronger effects mean less misclassification. In this model U is an unchanging binary variable that represents all other causes of oxidative stress present during the development of adenoma. Approximately 30% of the population is assumed to have U = 1, and U is assumed to have equal effects on both $Oxstress_2$ and $Oxstress_3$. The effects of U on oxidative stress created a correlation between $Oxstress_2$ and $Oxstress_3$; $Oxstress_2$ is measured with error, since $Oxstress_3$ is used as a proxy. As with BMI, if U has strong effects on both oxidative stress measurements, the measurement error will be less pronounced.

In the simulations, we considered different magnitudes of the effects of U on the Oxstress measurements, and of BMI_1 on BMI_2, and BMI_2 on BMI_3 to represent greater amounts of measurement error/misclassification. These baseline effects were chosen to correspond with bivariate correlation coefficients (ρ) of approximately 0.90 between any two variables, and then lower values of 0.80, 0.70, 0.60, and 0.50 were considered, as noted in Table 4.1. Five hundred populations of 1,000 subjects were simulated for each combination of parameters considered. To analyze the simulated data, we modeled an effect of BMI_3 on $Oxstress_3$ and an effect of $Oxstress_3$ on $Adenoma_3$ (representing the cross-sectional study design). Information bias was estimated by comparing the average estimated parameters for the 500 populations to the true population parameters.

The degree of bias for the effects of BMI on oxidative stress, and oxidative stress on colorectal adenoma are shown in Table 4.1. As the correlation between paired variables weakens, a larger amount of information bias is observed. The estimated effect of oxidative stress on adenoma decreased from an odds ratio (OR) of 1.84 to 1.40 when the correlation between $Oxstress_2$ and $Oxstress_3$ decreased from 0.90 to 0.50. Similarly, the estimated effect of BMI on

oxidative stress decreased from a 1.13 s.d. change in oxidative stress to 0.78 when the correlations across BMI decreased in the same fashion. The results from Table 4.1 show bias towards the null for both estimated effects.

Alternative model – reverse causation

If a health outcome developed before a biologic phenomenon (e.g., oxidative stress) was measured, the measured phenomenon could not have been a cause. However, depending on the conditions, the health outcome could have been caused by the same biologic phenomenon at a previous time, but also affect later measurements of the phenomenon ("reverse causation"). In our example, colorectal adenoma may take years to develop, whereas the five markers of oxidative stress measure oxidation over a comparatively shorter, more recent time frame. However, because of the cross-sectional study design the biomarker measurements could conceivably have been affected, in part, by the adenoma.

Reverse causation between oxidative stress and colorectal adenoma is represented in Figure 4.3. As in Figure 4.2, U represents all external causes of oxidative stress and has equal effects on $Oxstress_1$ and $Oxstress_2$ so that they are correlated ($\rho_{Oxstress}$), and BMI_1 is a cause of BMI_2 (ρ_{BMI}). We also assumed that BMI_1 is a cause of $Oxstress_1$ and that BMI_2 is a cause of $Oxstress_2$ (in the cross-sectional study, all variables are measured at time 2). Using the same parameters and simulation features as in the true causal model, we investigated the degree of potential bias due to reverse causation. The correlation between $Oxstress_1$ and $Oxstress_2$ was set as in the primary model, but before adenoma affected $Oxstress_2$. Therefore, $\rho_{Oxstress}$ refers to the correlation between $Oxstress_1$ and $Oxstress_2$ if adenoma had had no effect on $Oxstress_2$.

The degree of bias due to reverse causation is summarized in Table 4.2 and Figure 4.4. When the effect of adenoma increased $Oxstress_2$ by 1.25 s.d. (β), the estimated effect of BMI on oxidative stress and that of oxidative stress on colorectal adenoma were both biased away from

the null. The bias in the estimated effect of BMI on oxidative stress was moderate and did not vary greatly in either scenario: holding the BMI correlation constant at 0.90 while the oxidative stress correlation varied (range 1.37-1.42 s.d.), and holding oxidative stress correlation at 0.90 while BMI correlation varied (range 1.34-1.37 s.d.). In contrast to the effect of BMI on $Oxstress_2$, the bias in the estimated effect of $Oxstress_2$ on adenoma increased as the correlation became weaker. This result suggests that the degree of bias in the estimated effect of oxidative stress on colorectal adenoma depends on the effect of adenoma on oxidative stress, and on the association between the two oxidative stress measurements. The biased effect estimate as a function of these two properties is depicted in Figure 4.4. As is shown in the figure, when the correlation between the two oxidative stress measurements is strong (0.90), the bias was small regardless of the effect of adenoma on $Oxstress_2$. However, when the correlation was weaker (0.80 and 0.70), the bias was stronger, especially when the effect of adenoma on $Oxstress_2$ increased.

In summary, the temporal and causal assumptions required in SEMs are important. Cross-sectional studies are likely to have a degree of measurement error due, in part, to the use of concurrently measured biomarkers as proxies for earlier, causally relevant values. However, if reverse causation is not an issue, bias is likely towards the null. The magnitude of bias depends on the correlation between the measured variables and their earlier, causally relevant counterparts. In studies where reverse causation is a possibility, as hypothetically illustrated in our oxidative stress and adenoma example, the direction of bias can be away from the null. When reverse causation is an issue, the magnitude of bias also depends on the strength of the effect of the outcome on the measured intermediate. Longitudinal study designs can overcome some of these limitations by assuring that measured values precede their putative effects. These types of studies can also characterize how biomarkers might change during the development and progression of disease. If repeated measurements of biomarkers are not feasible, simple follow-

up studies with measurement of biomarker levels at baseline, with adequate follow-up, can provide evidence concerning the temporal relationships between biomarkers and outcome.

SELECTION BIAS

Selection bias can affect essentially all epidemiologic studies, both observational (cross-sectional, case-control, and follow-up) and experimental (clinical trials). Exclusion criteria, competing risks, and selection based on disease-free survival can all create bias. For observational studies of incidence, the usual approach is to exclude subjects who were not at risk for the outcome at the onset of the study (157); however, doing so can also create bias. Suppose that the exposure does have an effect on the outcome. Then selecting subjects who are disease-free at the beginning of the study inherently conditions on an effect of the exposure and may create a selection bias if there are any unmeasured risk factors for the outcome (158). This bias is illustrated in studies of obesity, end-stage renal disease (ESRD), and mortality, where, conditional on ESRD (an effect of obesity), the association between obesity and mortality is the opposite of what is expected, which is frequently called "reverse epidemiology" (159-161). Despite the apparently paradoxical results, an examination of the causal relationships between variables reveals that "reverse epidemiology" is likely due to collider bias (structural selection bias) (158, 161, 162). Investigations into this type of bias suggest that it is expected to be influential under the non-null conditions unless all risk factors for an outcome are adjusted, a nearly impossible situation (162).

SEMs used to investigate biologic constructs are vulnerable to collider bias, much like conventional analyses of observational studies. However, SEMs often involve some form of mediation analysis (i.e., estimation of direct and indirect effects) and under certain circumstances, this bias may manifest itself differently than in traditional regression analyses. We present bias from this source using adenoma, oxidative stress, and a generic exposure as a case example.

Figure 4.5 represents the causal relationships between an exposure (binary), oxidative stress (continuous), an uncontrolled harmful risk factor U (binary), and colorectal adenoma (binary). To simplify the model we assumed the temporal and causal relationships as indicated in Figure 4.5, and the absence of measurement error or misclassification. Given the model as summarized in Figure 4.5, DAG theory suggests a biasing path from exposure to adenoma, through survival and the U factor, when survival is conditioned on at the beginning of the study (82).

The SEM that incorporates the causal relationships in Figure 4.5 was considered to be the true causal model. We then simulated selection bias, due to conditioning on survival. In the simulation, the causal effects of interest were exposure on oxidative stress, exposure on adenoma, and oxidative stress on adenoma, and were set according to -0.634 s.d. change, a 1.42 OR, and a 1.98 OR, respectively (also shown in Figure 4.5). The causal effects of exposure on survival, of U on survival, and of U on adenoma were allowed to vary to illustrate the magnitude of selection bias under these different scenarios. We considered five different odds ratio values for the effect of exposure on survival (β = 2.0, 3.5, 5.0, 7.5, and 10.0). For each odds ratio, the effect of U on survival was taken to be the inverse of that OR ($1/\beta$ = 0.91, 0.50, 0.29, 0.20, 0.13, 0.10). We considered three different odds ratios for the effect of U on adenoma (γ = 1.1, 2.0, 3.5) and, additionally, considered three different values for the proportion of subjects who survived disease-free to the start of the study (survival = 0.90, 0.75, 0.60). The probability of being exposed was set at 50%, identical to the probability of being exposed to U.

For each combination of parameters, 500 populations were simulated with 1,000 subjects each. The simulated data were analyzed with a model, selecting (conditioning) subjects who survived to the beginning of the study (survival=1); we did not include the U variable in the SEM so as to represent an unmeasured risk factor. We estimated the effects of exposure on oxidative

stress and colorectal adenoma and of oxidative stress on colorectal adenoma, averaged across all 500 populations.

The potential degree of bias due to conditioning on disease-free survival is displayed in Table 4.3. In no scenario are the effects of oxidative stress on adenoma biased, and neither are the effects of exposure on oxidative stress. Selection bias for the scenarios considered affected only the estimate of exposure on adenoma, and the bias was not particularly strong under most conditions. Under a baseline survival probability of 0.90, in most scenarios the bias was less than 10%. Only when exposure had a strong effect on survival (OR $\geq$ 7.5), and U had strong effects on both adenoma (OR $\geq$ 2.0) and survival (OR $\leq$ 0.13), were effect estimates biased more than 10%. When the survival probability was lower (0.75 or 0.60), the potential bias was larger, but was still relatively modest unless strong effects on survival and adenoma were assumed.

In summary, the potential for selection bias is an important consideration for essentially all studies, including those investigating a biologic construct using SEMs. However, for the scenarios considered, the bias was small, unless the causal effects on the collider and on the pathway to adenoma were strong. While selection bias can present in different scenarios (e.g., differential loss to follow-up, sampling bias, participation bias), each one depicted in a DAG involves conditioning on a collider variable, thereby inducing a spurious association between the causes of the collider (158). The overall conclusions, that the bias is dependent on its underlying causal effects, are transferable to other situations of selection bias.

CONFOUNDING

Confounding is defined as the mixing of effects of an extraneous variable with those of the factor of interest so as to distort the observed result (157). In DAG theory, confounding is often depicted as a variable that causes both the exposure of interest and the outcome of concern,

thereby creating a confounding path (82). Even though confounding and a confounding path are not always the same, they are closely related (163).

Potential influence of confounding on an effect estimate is well recognized in epidemiology (157). In observational studies where the goal is to estimate an effect, unmeasured confounding is a common concern. However, when an exposure is properly randomized, with a large population, the effect estimate is expected to not be confounded (157). For the SEM analyses, however, confounding may still be a concern, even in a randomized study if the unmeasured confounder influences the effects to and from the intermediate variable (29-34, 148).

A DAG reflecting the causal relationships involving an exposure (binary), an intermediate oxidative stress variable (continuous), a binary unmeasured confounder (U) influencing the effect of oxidative stress on the outcome, and colorectal adenoma (binary outcome) is shown in Figure 4.6. This model assumes that the exposure is a randomized intervention with 50% exposed, and therefore there should be no confounding of the total (overall) effect of the exposure on adenoma. Additionally, we assumed temporal and causal relationships, and no measurement error or selection bias.

To simulate confounding, oxidative stress was assumed to increase the risk odds of adenoma two-fold (OR = 2.0) for each s.d. increase, and the OR for exposure was assumed to be 0.75. The exposure was assumed to decrease oxidative stress by -0.50 s.d. (also shown in Figure 4.6). In the scenarios considered, we evaluated the magnitude of confounding for different effects of U on oxidative stress (β = 0.33, 0.66, or 1.0 s.d.), for different effects of U on adenoma (OR = 1.5, 2.0, 2.5), and for different frequencies of exposure to U (10%, 25%, 50%). For each combination of parameters, 500 populations were simulated with 1,000 subjects each. In the SEM used to analyze the simulated data, we did not include the U variable so that confounding remained uncontrolled. We averaged the estimated effects of exposure on oxidative stress and

adenoma and of oxidative stress on adenoma from the 500 populations. Comparing the estimated effects with the known population parameters provided a measure of potential confounding for each scenario.

The magnitude of potential confounding due to an uncontrolled confounder, for the effects of exposure on oxidative stress and adenoma is displayed in Table 4.4. In no scenario was the effect of exposure on oxidative stress biased, and remained unchanged at -0.50 s.d. However, the effect of oxidative stress on adenoma, and the effect of exposure on adenoma were both biased. When the effect of U on oxidative stress was weak (0.33 s.d.), or when the effect of U on adenoma was weak (OR = 1.5), the magnitude of error in the observed effect of oxidative stress on adenoma was also weak (range = 2.5-8.5%). The bias was strongest (19%) when U had a strong effect on both adenoma (OR = 2.5) and oxidative stress (1.0 s.d.). The confounding error in the direct effect of exposure on adenoma was also present but was not as strong (range 1.3 - 9.3%) as that for the corresponding effect of oxidative stress. In scenarios with different proportions of subjects that were exposed to U, the observed magnitude of confounding differed slightly.

In summary, the magnitude of confounding in SEM-based analyses (regardless of the study design) depends on the effects of the uncontrolled confounder on both the intermediate and the outcome variable. For confounding to meaningfully change an estimate of interest (> 10% error), these effects have to be strong. Less important is the proportion of study subjects exposed to the unmeasured confounding factor.

DISCUSSION

Valid estimation of effects involves making correct assumptions and avoiding biases. These considerations apply not only to traditional epidemiologic studies, but also to those analyzed using SEMs. SEMs differ from traditional regression models both in terms of the

quantity and the quality of required assumptions, and in terms of the expected manifestations of biases. In this communication, we have explicated some of the key assumptions and quantified some of the potential biases that are specific to SEM-based analyses.

SEMs with latent variables allows for the estimation of biologic constructs by modeling the shared covariance of multiple measures (including biomarkers). This is a very powerful analytical tool, but one that requires many additional assumptions for valid estimation of effects. Not surprisingly, many of those assumptions center on the latent variable and its components. Important issues to consider include whether the shared covariance involving biomarkers or other measures is exclusively caused by the construct of interest, or whether they share an additional common cause. Another important issue is whether exposures of interest can affect biomarkers, or co-vary with them, independent of the construct; the same issue arises with biomarkers and the outcome. For all of the above reasons, the best guide to using SEMs is to start with a well-defined underlying biologic theory. In cancer research, cellular and molecular biology can provide the necessary empirical and theoretical evidence surrounding the biologic phenomenon to guide a SEM.

Not unlike traditional analytic methods, SEMs must deal with potential information, selection, and confounding biases. In cross-sectional studies, particular attention should be paid to the possibility of reverse causation. If the outcome can meaningfully affect the biomarker measurements, longitudinal or follow-up study designs remain the only option. Selection bias, due to conditioning on disease-free survival, is another potential issue (162). However, the magnitude of the induced bias depends strongly on the effect of the unmeasured factor on both the outcome and the likelihood of survival. Potential confounding involving an intermediate is an important consideration when analyzing effects using a SEM. Even under a randomized study design, unmeasured confounding of the intermediate's effect on the outcome may introduce bias.

Careful consideration of all potential confounders, for exposure and intermediate, is required in an SEM.

Using the previous analyses in chapters 2 and 3 as examples allowed a semi-quantitative evaluation of the validity of those results and conclusions. The assumptions that underscore the latent variable in both chapters are strong, as oxidative stress mediates all causes to the biomarkers. If those assumptions are violated, then the latent variable could represent another biologic mechanism in addition to oxidative stress. Given that CRP is an acute inflammation protein, violation of the assumptions could result in a latent variable consisting of oxidative stress and inflammation, a limitation we acknowledged in those chapters. For biases, all could potentially affect our results. The simulations in this chapter suggest that potential biases from selection and unmeasured confounding are not likely strong enough to refute our overall conclusions. However, reverse causation is a real concern as it could have biased our chapter 3 oxidative stress on colorectal adenoma estimate (OR = 2.61) away from the null. This is also a recognized limitation in that chapter and future studies should consider follow-up designs to preclude this potential bias.

SEMs may provide unique research opportunities for studies of biologic phenomena. However, there are particular concerns regarding potential biases and causal assumptions in the application of this method. Some may be alleviated or mitigated by study design, but all require sound biologic theory.

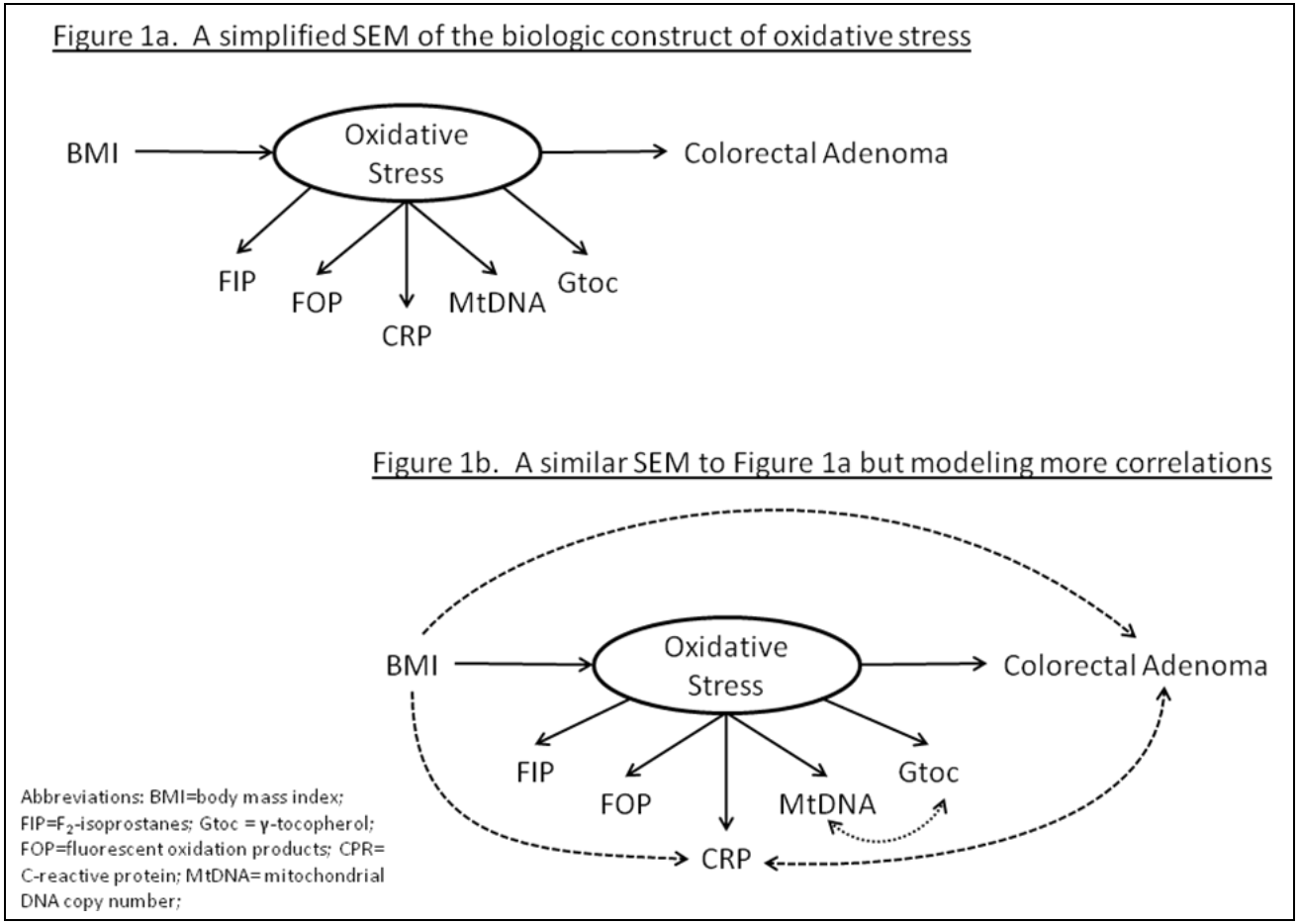

Figure 4.1. SEMs demonstrating common causal assumptions in studies of biologic phenomena

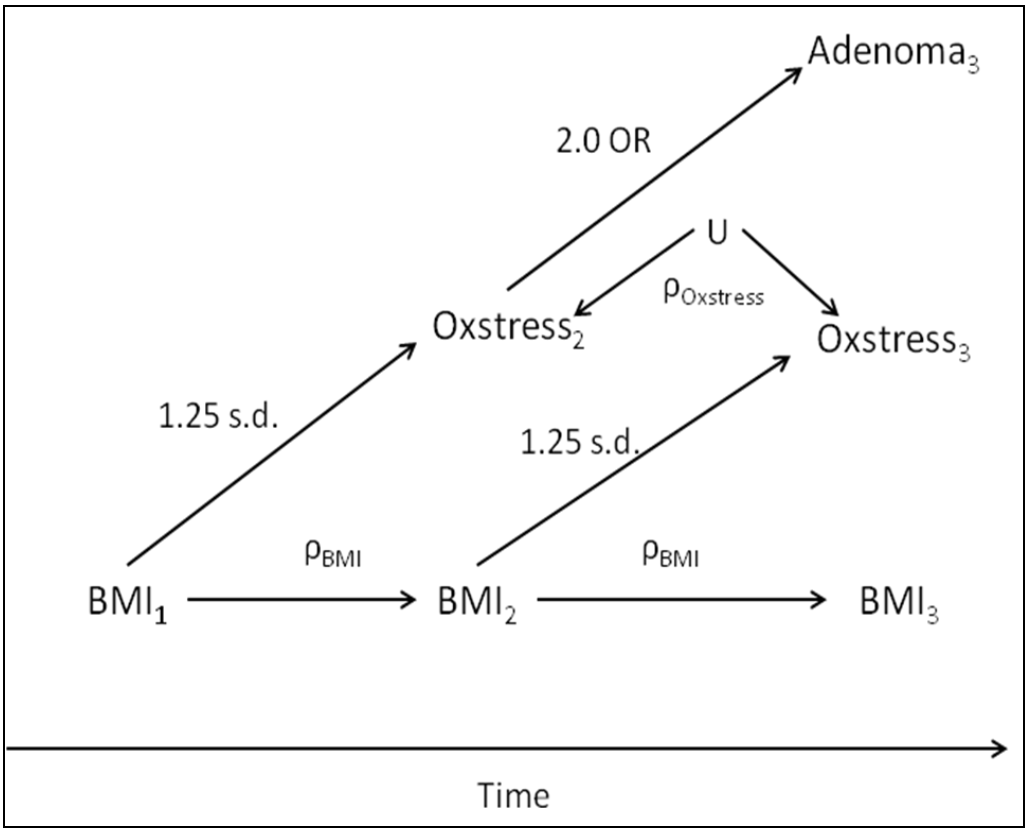

Figure 4.2. Model of measurement error/misclassification of body mass index (BMI) and oxidative stress (Oxstress) over time

Table 4.1. Estimated biased effects of BMI on oxidative stress, and oxidative stress on colorectal adenoma from measurement error/misclassification from Figure 4.2

		Correlation between $Oxstress_2$ and $Oxstress_3$ holding BMI correlations at 0.90				
Causal Effects	Population	$\rho_{Oxstress}$ = 0.9	0.80	0.70	0.60	0.50
BMI -> Oxstress	1.25 s.d.	1.13	1.13	1.13	1.13	1.13
Oxstress -> Adenoma	2.00 OR	1.84	1.73	1.57	1.49	1.40
		Correlation between BMI_1 and BMI_2, BMI_2 and BMI_3 holding Oxstress correlations at 0.90				
	Population	ρ_{BMI} = 0.9	0.80	0.70	0.60	0.50
BMI -> Oxstress	1.25 s.d.	1.13	1.04	0.93	0.88	0.78
Oxstress -> Adenoma	2.00 OR	1.84	1.84	1.83	1.83	1.83

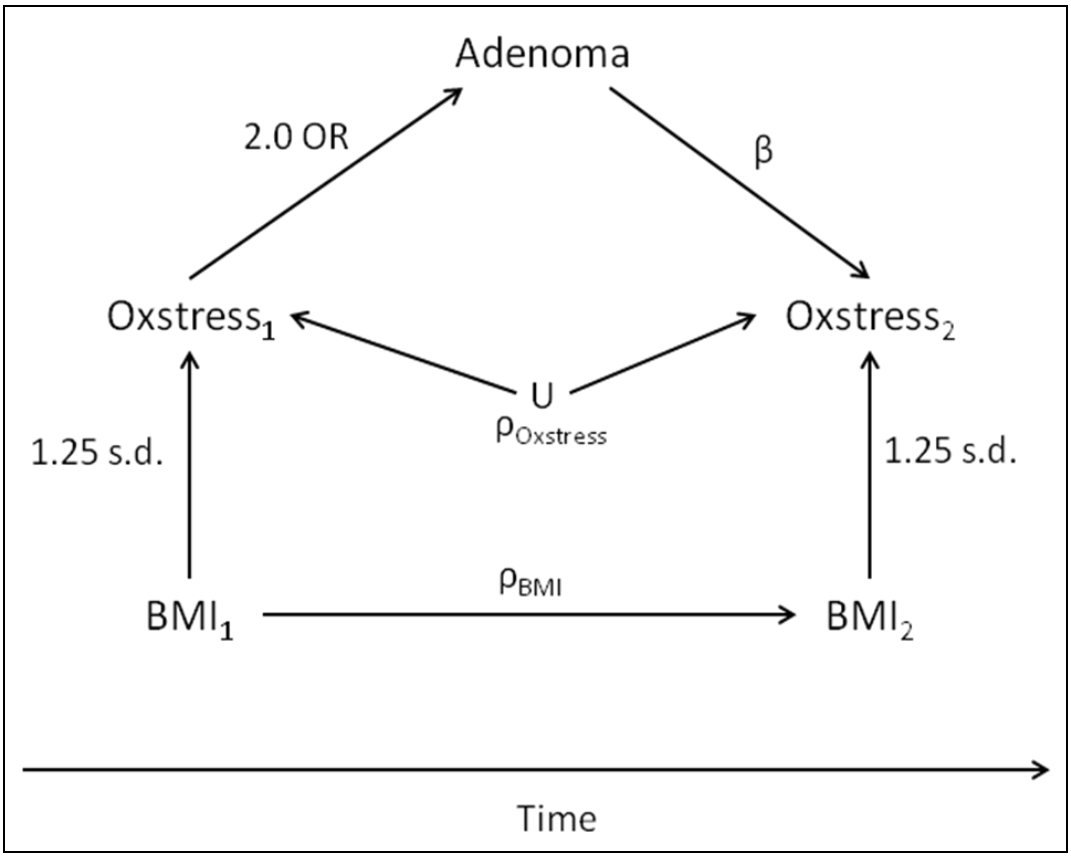

Figure 4.3. Model of reverse causation between adenoma and oxidative stress (Oxstress)

Table 4.2. Degree of bias from reverse causation when adenoma has a 1.25 standard deviation effect on oxidative stress from Figure 4.3

		Correlation between $Oxstress_1$ and $Oxstress_2$ holding BMI correlations at 0.90				
Causal Effects	Population	$\rho_{Oxstress}$ = 0.9	0.80	0.70	0.60	0.50
BMI -> Oxstress	1.25 s.d.	1.37	1.40	1.41	1.42	1.41
Oxstress -> Adenoma	2.00 OR	1.97	2.02	2.19	2.29	2.44
		Correlation between BMI_1 and BMI_2 holding Oxstress correlations at 0.90				
	Population	ρ_{BMI} = 0.9	0.80	0.70	0.60	0.50
BMI -> Oxstress	1.25 s.d.	1.37	1.37	1.35	1.35	1.34
Oxstress -> Adenoma	2.00 OR	1.97	1.96	1.96	1.95	1.95

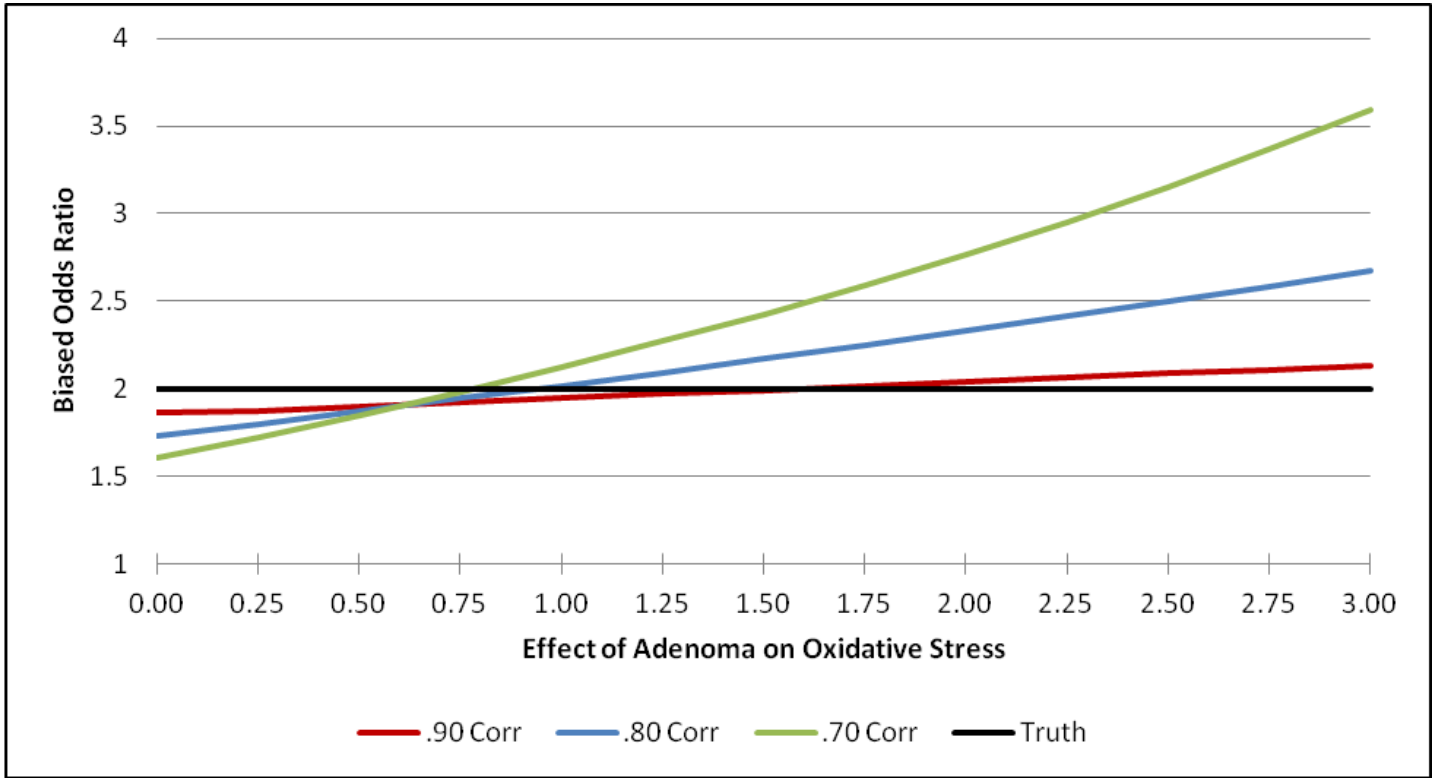

Figure 4.4. Degree of bias due to reverse causation as a function of adenoma's effect on oxidative stress and the correlation between measured oxidative stress and the causally relevant oxidative stress

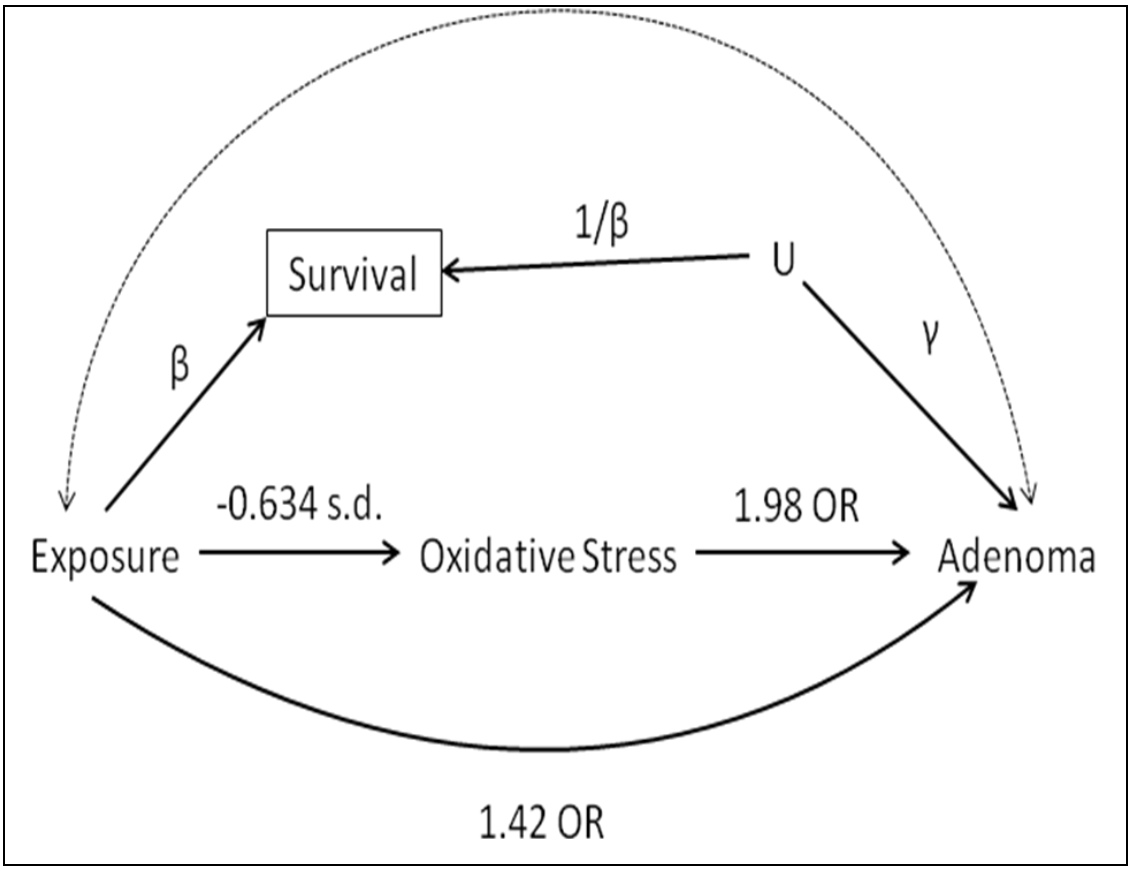

Figure 4.5. Model of selection bias through conditioning on disease-free survival with an unmeasured risk factor for adenoma

Table 4.3. Degree of bias from disease-free selection with an unmeasured risk factor for adenoma from Figure 4.5

		Baseline survival = 0.90					**Baseline survival = 0.75**					**Baseline survival = 0.60**				
		Odds ratio (OR) effects of exposure on survival and 1/OR effects of Unknown risk factor (U) on survival														
Effects	Causal	OR = 2.0	3.5	5.0	7.5	10.0	2.0	3.5	5.0	7.5	10.0	2.0	3.5	5.0	7.5	10.0
		Effect of U on adenoma is OR = 1.1					**Effect of U on adenoma is OR = 1.1**					**Effect of U on adenoma is OR = 1.1**				
Oxst->C.A.	1.98	2.00	2.00	1.99	2.00	2.00	1.99	2.00	2.00	1.99	1.99	1.99	1.99	1.99	1.99	1.99
Exp->C.A.	1.42	1.42	1.42	1.42	1.43	1.46	1.40	1.44	1.46	1.45	1.46	1.44	1.50	1.49	1.52	1.57
Exp->Oxst	-0.63	-0.63	-0.64	-0.64	-0.64	-0.63	-0.63	-0.63	-0.63	-0.63	-0.64	-0.63	-0.63	-0.63	-0.63	-0.63
		Effect of U on adenoma is OR = 2.0					**Effect of U on adenoma is OR = 2.0**					**Effect of U on adenoma is OR = 2.0**				
Oxst->C.A.	1.98	1.98	1.97	1.97	1.98	1.98	1.97	1.98	1.98	1.97	1.98	1.97	1.98	1.98	1.98	1.99
Exp->C.A.	1.42	1.43	1.45	1.48	1.52	1.56	1.43	1.48	1.54	1.59	1.64	1.42	1.51	1.56	1.67	1.78
Exp->Oxst	-0.63	-0.63	-0.64	-0.64	-0.64	-0.63	-0.63	-0.63	-0.63	-0.63	-0.64	-0.63	-0.63	-0.63	-0.63	-0.63
		Effect of U on adenoma is OR = 3.5					**Effect of U on adenoma is OR = 3.5**					**Effect of U on adenoma is OR = 3.5**				
Oxst->C.A.	1.98	1.98	1.98	1.98	1.98	1.98	1.97	1.97	1.97	1.97	1.98	1.96	1.96	1.98	1.98	1.98
Exp->C.A.	1.42	1.44	1.49	1.54	1.61	1.69	1.46	1.56	1.65	1.76	1.86	1.46	1.58	1.68	1.85	2.03
Exp->Oxst	-0.63	-0.63	-0.64	-0.64	-0.64	-0.63	-0.63	-0.63	-0.63	-0.63	-0.64	-0.63	-0.63	-0.63	-0.63	-0.63

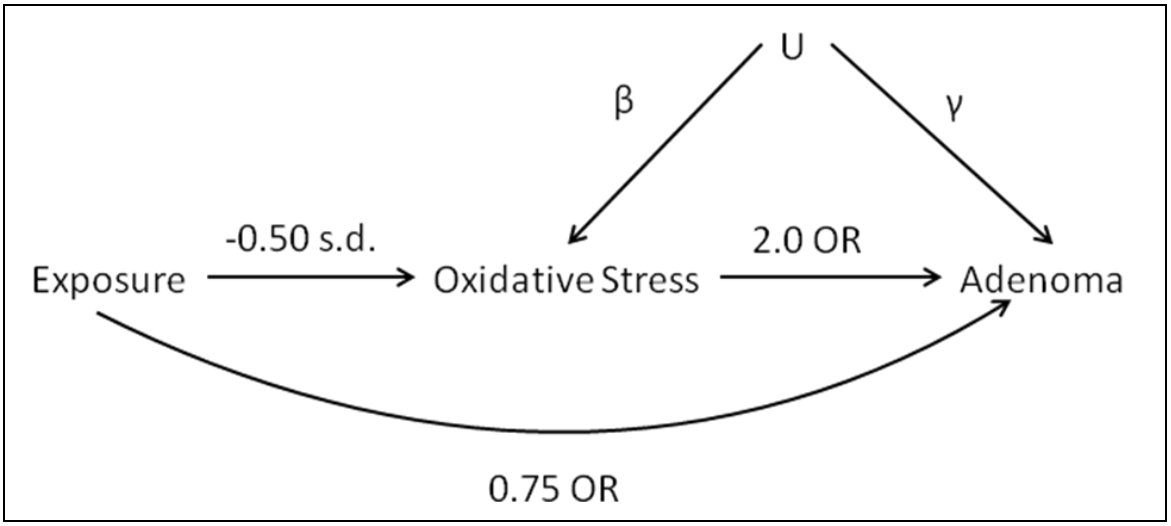

Figure 4.6. Model of unmeasured confounding of the effect of oxidative stress on adenoma

Table 4.4. Degree of bias from unmeasured confounder of the effect of oxidative stress on adenoma from Figure 4.6

		U on Oxstress = 0.33 s.d.			**U on Oxstress = 0.66**			**U on Oxstress = 1.0**		
		Changing odds ratios (OR) of U on Adenoma (1.5, 2.0, 3.5)								
Effects	Causal	OR = 1.5	2.0	2.5	1.5	2.0	2.5	1.5	2.0	2.5
					Probability of U exposed = 10%					
Exp -> OxStress	-0.50	-0.50	-0.50	-0.50	-0.50	-0.50	-0.50	-0.50	-0.50	-0.50
Exp -> Adenoma	0.75	0.77	0.77	0.78	0.78	0.79	0.79	0.78	0.79	0.82
Oxstress -> Adenoma	2.00	2.05	2.08	2.10	2.10	2.15	2.20	2.14	2.23	2.32
					Probability of U exposed = 25%					
Exp -> OxStress	-0.50	-0.50	-0.50	-0.50	-0.50	-0.50	-0.50	-0.50	-0.50	-0.50
Exp -> Adenoma	0.75	0.77	0.78	0.79	0.77	0.80	0.80	0.79	0.80	0.81
Oxstress -> Adenoma	2.00	2.07	2.12	2.14	2.13	2.20	2.27	2.17	2.29	2.38
					Probability of U exposed = 50%					
Exp -> OxStress	-0.50	-0.50	-0.50	-0.50	-0.50	-0.50	-0.50	-0.50	-0.50	-0.50
Exp -> Adenoma	0.75	0.76	0.78	0.79	0.77	0.80	0.81	0.79	0.81	0.81
Oxstress -> Adenoma	2.00	2.09	2.12	2.15	2.13	2.20	2.24	2.17	2.25	2.31

CHAPTER 5. DISCUSSION AND FUTURE DIRECTIONS

OVERVIEW OF FINDINGS

The overall goal of this dissertation was to assess the utility of structural equation models (SEMs) with latent variables to study oxidative stress, its determinants, and its association with colorectal adenoma. Specifically, I wanted to know whether a latent variable identified from of F_2-isoprostanes (FIP), fluorescent oxidation products (FOP), C-reactive protein (CRP), mitochondrial DNA (MtDNA) copy number, and γ-tocopherol (Gtoc), would provide a better measure of oxidative stress than could be provided via traditional analyses limited to observed variables. Additionally, I was interested in the validity of causal assumptions and the magnitude of potential biases pertaining to the SEM method. The objectives and goals of this dissertation were achieved through three distinct but related research projects.

In the first project, I investigated whether a SEM would suitably identify and characterize oxidative stress from the five *a priori* selected biomarkers (FIP, FOP, CRP, MtDNA copy number, and Gtoc). Using the cross-sectional Markers of Adenomatous Polyps (MAP) studies I and II, and the Study of Race Stress and Hypertension (SRSH), a latent variable was constructed, primarily from FIP, Gtoc, and CRP, suggesting a common cause of all three biomarkers. Additionally, the latent variable was statistically significantly associated with higher BMI and a history of smoking in the positive direction, and a history of drinking and regular aspirin use in the negative direction. The latent variable was also associated with several plasma antioxidant markers in the theorized directions. These results provided enough evidence to justifiably call the latent variable "oxidative stress". The associations of pro- and antioxidant exposures with oxidative stress were stronger than the respective associations with each biomarker individually. This observation demonstrates that the latent variable provides additional information not

obtainable by any single biomarker. The results of the study also show that the SEM method may provide a better measurement of oxidative stress than traditional regression methods.

In the second project, I examined the association between the latent oxidative stress variable and newly diagnosed colorectal adenoma. The construction of the latent variable was similar to that in the first project. In the SEM analyses, higher oxidative stress was strongly, statistically significantly associated with higher odds of colorectal adenoma after adjusting for potential confounders. Moreover, higher BMI, a history of drinking, and regular aspirin use, were indirectly associated with colorectal adenoma mediated by oxidative stress. The results of the second project demonstrate that the SEM method may produce a stronger association between oxidative stress and colorectal adenoma than what was reported in the literature.

In the third project, using the previous sections of this dissertation as case examples, I critically evaluated the causal assumptions of SEMs when applied to studies of biologic phenomena. Fundamental SEM assumptions involving the latent variable were explicated specifically with respect to potential systematic error introduced by information bias, conditional selection into the study (a.k.a. collider bias), and confounding was quantitatively assessed through Monte Carlo model simulations. The results from the simulation analyses indicated that the magnitude of error from confounding or collider bias, which can be manifested differently in SEMs than in traditional regression models, is likely to be modest unless unadjusted factors have very strong effects on other variables. The main concern with information bias in cross-sectional studies such as those used in this dissertation is reverse causation. In our example, if adenoma did not affect concurrently measured levels of oxidative stress, then the bias was towards the null. Otherwise the bias is expected to be away from the null, and could be quite strong, depending on other factors. It appears that SEMs require strong assumptions but may offer unique research perspectives for studies of biologic phenomena.

Overall, the results from the three projects were in support of the primary hypothesis that a latent variable composed of FIP, FOP, CRP, MtDNA copy number, and Gtoc, would be a valid measurement of oxidative stress. The use of SEM with latent variables adds to the existing literature on oxidative stress by offering insights that cannot be obtained from more traditional regression analyses. The latent oxidative stress variable was more strongly associated with a health outcome and with pro- and antioxidant exposures than any of its individual markers, and while the method makes strong assumptions, the potential biases were unlikely to completely explain the observed results.

Implications and Future Directions

Although SEMs are common in the fields of psychology, sociology, and economics, they are infrequently used in epidemiology (164). The application of SEMs to study biologic processes and phenomena is rare (165), and to our knowledge the studies in this dissertation are the first to use SEMs to investigate oxidative stress.

For measurement of oxidative stress, the next logical step is to include more biomarkers of oxidation. Consideration of the findings in this dissertation should guide the additional biomarkers. Adding 8-oxo-7,8-dihydro-2'-deoxyguanosine could enable the latent variable to better measure DNA damage, something we were unable to do with MtDNA copy number. Malondialdehyde or another lipid peroxidation biomarker could be used together with FIP for better identification of lipid peroxidation. Oxidized proteins are a significant by-product of oxidative stress, and the addition of protein carbonyls could facilitate a more complete latent measure of oxidative stress. Also, including inflammation biomarkers, such as interleukin 6 or tumor necrosis factor alpha, may allow for the identification of an oxidative stress and inflammation latent variable. It is possible that with enough markers a SEM can be used to reveal multiple constructs that are inter-related in a network of biologic phenomena. Before

constructing a more complex SEM, however, it is important to keep in mind that as the model is expanded the number of required assumptions becomes larger. These additional assumptions would have to be backed by strong biologic theory.

Other future areas of research regarding a latent variable of oxidative stress include using different explanatory variables or stratifying populations. In this dissertation, the exposure variables were primarily questionnaire-based lifestyle factors but included some plasma antioxidant variables. Although oxidative stress is affected by exogenous stimuli, it is also affected by endogenous enzymatic mechanisms (1). Heritable genetic factors could influence oxidative stress through the modification of enzymes. Future research of potential genetic alterations could find important associations with a latent oxidative stress variable. Another area of research is constructing sex- or race-/ethnicity-specific latent oxidative stress variables. In this dissertation, in which I used relatively small datasets, the loading factors for the latent variable were only partially invariant in men and women and across races/ethnicities. It is possible that inherent biologic differences between men and women or across racial/ethnic groups affect the measurement of oxidative stress, necessitating stratification of the data in future studies with larger sample sizes.

In the second study of this dissertation oxidative stress was strongly associated with colorectal adenoma. This finding should be interpreted with caution because of the cross-sectional study design. It is possible that the measurements of oxidative stress were not causally relevant to the development of adenoma. For future research follow-up studies with oxidation measurements assayed at study initiation should be considered. Longitudinal study designs with multiple oxidative stress measurements over time would be more suitable for investigating changes in oxidative stress during the development of adenoma; however, such studies may not be feasible in large populations. Since adenoma is a precursor to carcinoma, the next logical step would be a similar study of colorectal cancer. However, colorectal carcinoma is more likely to

exert systemic effects, including an increase in oxidative stress, and for this reason, reverse causation becomes a particularly relevant concern. If the goal is to estimate the effect of oxidative stress on colorectal cancer, a follow-up study becomes even more necessary.

Basic science and clinical evidence indicate that oxidative stress can affect the risk of many different age-related conditions including several types of cancer and cardiovascular disease (7). Thus, future research directions may include application of SEM with latent variables to investigate the role of oxidative stress in other health outcomes as well as, more generally, aging and longevity.

References

1. Jones DP, DeLong MJ. Detoxification and Protective Functions of Nutrients. In: Stipanuk M, editor. Biochemical, Phsyiological and Molecular Aspects of Human Nutrition. 1st ed: Saunders Elsevier; 2000. p. 901-16.
2. Halliwell B. Free Radicals and Other Reactive Species in Disease. eLS: John Wiley & Sons, Ltd; 2001.
3. Valko M, Rhodes CJ, Moncol J, Izakovic M, Mazur M. Free radicals, metals and antioxidants in oxidative stress-induced cancer. Chem Biol Interact. 2006 Mar 10;160(1):1-40.
4. Halliwell B. Oxidative stress and cancer: have we moved forward? Biochem J. 2007 Jan 1;401(1):1-11.
5. Sies H. Oxidative stress. London: Orlando : Academic Press; 1985.
6. Jones DP. Redefining oxidative stress. Antioxid Redox Signal. 2006 Sep-Oct;8(9-10):1865-79.
7. Giustarini D, Dalle-Donne I, Tsikas D, Rossi R. Oxidative stress and human diseases: Origin, link, measurement, mechanisms, and biomarkers. Crit Rev Clin Lab Sci. 2009;46(5-6):241-81.
8. Ashok BT, Ali R. The aging paradox: free radical theory of aging. Exp Gerontol. 1999 Jun;34(3):293-303.
9. Goodman M, Bostick RM, Kucuk O, Jones DP. Clinical trials of antioxidants as cancer prevention agents: past, present, and future. Free Radic Biol Med. 2011 Sep 1;51(5):1068-84.
10. Wright ME, Mayne ST, Stolzenberg-Solomon RZ, Li Z, Pietinen P, Taylor PR, et al. Development of a comprehensive dietary antioxidant index and application to lung cancer risk in a cohort of male smokers. Am J Epidemiol. 2004 Jul 1;160(1):68-76.
11. Young IS, Woodside JV. Antioxidants in health and disease. J Clin Pathol. 2001 Mar;54(3):176-86.
12. Goodman M, Bostick RM, Dash C, Flanders WD, Mandel JS. Hypothesis: oxidative stress score as a combined measure of pro-oxidant and antioxidant exposures. Annals of epidemiology. 2007 May;17(5):394-9.
13. Van Hoydonck PG, Temme EH, Schouten EG. A dietary oxidative balance score of vitamin C, beta-carotene and iron intakes and mortality risk in male smoking Belgians. J Nutr. 2002 Apr;132(4):756-61.
14. Agalliu I, Kirsh VA, Kreiger N, Soskolne CL, Rohan TE. Oxidative balance score and risk of prostate cancer: results from a case-cohort study. Cancer Epidemiol. 2011 Aug;35(4):353-61.
15. Goodman M, Bostick RM, Dash C, Terry P, Flanders WD, Mandel J. A summary measure of pro- and anti-oxidant exposures and risk of incident, sporadic, colorectal adenomas. Cancer Causes Control. 2008 Dec;19(10):1051-64.
16. Goodman M, Bostick RM, Gross M, Thyagarajan B, Dash C, Flanders WD. Combined measure of pro- and anti-oxidant exposures in relation to prostate cancer and colorectal adenoma risk: an update. Annals of epidemiology. 2010 Dec;20(12):955-7.
17. Slattery ML, Lundgreen A, Welbourn B, Wolff RK, Corcoran C. Oxidative balance and colon and rectal cancer: interaction of lifestyle factors and genes. Mutat Res. 2012 Jun 1;734(1-2):30-40.

18. Kong SY, Bostick RM, Flanders WD, McClellan WM, Thyagarajan B, Gross MD, et al. Oxidative balance score, colorectal adenoma, and markers of oxidative stress and inflammation. Cancer Epidemiol Biomarkers Prev. 2014 Mar;23(3):545-54.
19. Bertram JS. Carotenoids and gene regulation. Nutr Rev. 1999 Jun;57(6):182-91.
20. Chew BP, Park JS. Carotenoid action on the immune response. J Nutr. 2004 Jan;134(1):257S-61S.
21. Heber D, Lu QY. Overview of mechanisms of action of lycopene. Exp Biol Med (Maywood). 2002 Nov;227(10):920-3.
22. Braven J, Bonker GJ, Fenner ML, Tonge BL. The mechanism of carcinogenesis by tobacco smoke. Some experimental observations and a hypothesis. Br J Cancer. 1967 Sep;21(3):623-33.
23. Seitz HK, Stickel F. Molecular mechanisms of alcohol-mediated carcinogenesis. Nat Rev Cancer. 2007 Aug;7(8):599-612.
24. Williams CS, Smalley W, DuBois RN. Aspirin use and potential mechanisms for colorectal cancer prevention. J Clin Invest. 1997 Sep 15;100(6):1325-9.
25. Steele VE, Hawk ET, Viner JL, Lubet RA. Mechanisms and applications of non-steroidal anti-inflammatory drugs in the chemoprevention of cancer. Mutat Res. 2003 Feb-Mar;523-524:137-44.
26. Bollen KA. Structural equations with latent variables. New York: Wiley; 1989.
27. Kline RB. Principles and practice of structural equation modeling. 2nd ed. New York: Guilford Press; 2005.
28. Beran TN, Violato C. Structural equation modeling in medical research: a primer. BMC Res Notes. 2010;3:267.
29. Pearl J. Direct and indirect effects. Proceedings of the Seventeenth conference on Uncertainty in artificial intelligence; Seattle, Washington. 2074073: Morgan Kaufmann Publishers Inc.; 2001. p. 411-20.
30. Robins JM, Greenland S. Identifiability and exchangeability for direct and indirect effects. Epidemiology. 1992 Mar;3(2):143-55.
31. Cole SR, Hernan MA. Fallibility in estimating direct effects. Int J Epidemiol. 2002 Feb;31(1):163-5.
32. VanderWeele TJ. Bias formulas for sensitivity analysis for direct and indirect effects. Epidemiology. 2010 Jul;21(4):540-51.
33. Cai Z, Kuroki M, Pearl J, Tian J. Bounds on direct effects in the presence of confounded intermediate variables. Biometrics. 2008 Sep;64(3):695-701.
34. Petersen ML, Sinisi SE, van der Laan MJ. Estimation of direct causal effects. Epidemiology. 2006 May;17(3):276-84.
35. Pearl J. Causality: Models, Reasoning, and Inference. 2nd ed. New York: Cambridge University Press; 2009.
36. Kohen R, Nyska A. Oxidation of biological systems: oxidative stress phenomena, antioxidants, redox reactions, and methods for their quantification. Toxicol Pathol. 2002 Nov-Dec;30(6):620-50.
37. Montuschi P, Barnes PJ, Roberts LJ, 2nd. Isoprostanes: markers and mediators of oxidative stress. FASEB J. 2004 Dec;18(15):1791-800.
38. Wu T, Willett WC, Rifai N, Rimm EB. Plasma fluorescent oxidation products as potential markers of oxidative stress for epidemiologic studies. Am J Epidemiol. 2007 Sep 1;166(5):552-60.
39. Wu T, Rifai N, Roberts LJ, 2nd, Willett WC, Rimm EB. Stability of measurements of biomarkers of oxidative stress in blood over 36 hours. Cancer Epidemiol Biomarkers Prev. 2004 Aug;13(8):1399-402.

40. Basu S. F2-isoprostanes in human health and diseases: from molecular mechanisms to clinical implications. Antioxid Redox Signal. 2008 Aug;10(8):1405-34.
41. Siamakpour-Reihani S, Scarbrough PM, Wang F, Spasojevic I, Base K, Sedjo R, et al. Systemic markers of oxidative status and colorectal adenomatous polyps. Annals of epidemiology. 2012 Aug;22(8):587-91.
42. Kong SYJ, Bostick R, Flanders WD, McClellan W, Thyagarajan B, Gross M, et al. Oxidative Balance Score, Colorectal Adenoma and Markers of Oxidative Stress and Inflammation. Atlanta: Emory University; 2013.
43. Block G, Dietrich M, Norkus EP, Morrow JD, Hudes M, Caan B, et al. Factors associated with oxidative stress in human populations. Am J Epidemiol. 2002 Aug 1;156(3):274-85.
44. Fowke JH, Morrow JD, Motley S, Bostick RM, Ness RM. Brassica vegetable consumption reduces urinary F2-isoprostane levels independent of micronutrient intake. Carcinogenesis. 2006 Oct;27(10):2096-102.
45. Hopkins MH, Fedirko V, Jones DP, Terry PD, Bostick RM. Antioxidant micronutrients and biomarkers of oxidative stress and inflammation in colorectal adenoma patients: results from a randomized, controlled clinical trial. Cancer Epidemiol Biomarkers Prev. 2010 Mar;19(3):850-8.
46. Frankel EN. Lipid oxidation: Oily Press; 2005.
47. Dillard CJ, Tappel AL. Fluorescent damage products of lipid peroxidation. Methods Enzymol. 1984;105:337-41.
48. Frankel EN, Neff WE, Brooks DD, Fujimoto K. Fluorescence formation from the interaction of DNA with lipid oxidation degradation products. Biochim Biophys Acta. 1987 Jun 23;919(3):239-44.
49. Carew JS, Huang P. Mitochondrial defects in cancer. Mol Cancer. 2002 Dec 9;1:9.
50. Chatterjee A, Mambo E, Sidransky D. Mitochondrial DNA mutations in human cancer. Oncogene. 2006 Aug 7;25(34):4663-74.
51. Warburg O. On the origin of cancer cells. Science. 1956 Feb 24;123(3191):309-14.
52. Penta JS, Johnson FM, Wachsman JT, Copeland WC. Mitochondrial DNA in human malignancy. Mutat Res. 2001 May;488(2):119-33.
53. Hibi K, Nakayama H, Yamazaki T, Takase T, Taguchi M, Kasai Y, et al. Detection of mitochondrial DNA alterations in primary tumors and corresponding serum of colorectal cancer patients. Int J Cancer. 2001 Nov 1;94(3):429-31.
54. Chang SC, Lin PC, Yang SH, Wang HS, Liang WY, Lin JK. Mitochondrial D-loop mutation is a common event in colorectal cancers with p53 mutations. Int J Colorectal Dis. 2009 Jun;24(6):623-8.
55. Thyagarajan B, Wang R, Barcelo H, Koh WP, Yuan JM. Mitochondrial copy number is associated with colorectal cancer risk. Cancer Epidemiol Biomarkers Prev. 2012 Sep;21(9):1574-81.
56. Wang YC, Lee WC, Liao SC, Lee LC, Su YJ, Lee CT, et al. Mitochondrial DNA copy number correlates with oxidative stress and predicts mortality in nondiabetic hemodialysis patients. J Nephrol. 2011 May-Jun;24(3):351-8.
57. Jiang Q, Christen S, Shigenaga MK, Ames BN. gamma-tocopherol, the major form of vitamin E in the US diet, deserves more attention. Am J Clin Nutr. 2001 Dec;74(6):714-22.
58. Wu JH, Ward NC, Indrawan AP, Almeida CA, Hodgson JM, Proudfoot JM, et al. Effects of alpha-tocopherol and mixed tocopherol supplementation on markers of oxidative stress and inflammation in type 2 diabetes. Clin Chem. 2007 Mar;53(3):511-9.
59. Devaraj S, Leonard S, Traber MG, Jialal I. Gamma-tocopherol supplementation alone and in combination with alpha-tocopherol alters biomarkers of oxidative stress and inflammation in subjects with metabolic syndrome. Free Radic Biol Med. 2008 Mar 15;44(6):1203-8.

60. Himmelfarb J, Kane J, McMonagle E, Zaltas E, Bobzin S, Boddupalli S, et al. Alpha and gamma tocopherol metabolism in healthy subjects and patients with end-stage renal disease. Kidney Int. 2003 Sep;64(3):978-91.
61. Himmelfarb J, Phinney S, Ikizler TA, Kane J, McMonagle E, Miller G. Gamma-tocopherol and docosahexaenoic acid decrease inflammation in dialysis patients. J Ren Nutr. 2007 Sep;17(5):296-304.
62. Cooney RV, Franke AA, Wilkens LR, Gill J, Kolonel LN. Elevated plasma gamma-tocopherol and decreased alpha-tocopherol in men are associated with inflammatory markers and decreased plasma 25-OH vitamin D. Nutr Cancer. 2008;60 Suppl 1:21-9.
63. Ingles SA, Bird CL, Shikany JM, Frankl HD, Lee ER, Haile RW. Plasma tocopherol and prevalence of colorectal adenomas in a multiethnic population. Cancer Res. 1998 Feb 15;58(4):661-6.
64. Coussens LM, Werb Z. Inflammation and cancer. Nature. 2002 Dec 19-26;420(6917):860-7.
65. Reuter S, Gupta SC, Chaturvedi MM, Aggarwal BB. Oxidative stress, inflammation, and cancer: how are they linked? Free Radic Biol Med. 2010 Dec 1;49(11):1603-16.
66. Hulthe J, Fagerberg B. Circulating oxidized LDL is associated with subclinical atherosclerosis development and inflammatory cytokines (AIR Study). Arterioscler Thromb Vasc Biol. 2002 Jul 1;22(7):1162-7.
67. Yasunari K, Maeda K, Nakamura M, Yoshikawa J. Oxidative stress in leukocytes is a possible link between blood pressure, blood glucose, and C-reacting protein. Hypertension. 2002 Mar 1;39(3):777-80.
68. Abramson JL, Hooper WC, Jones DP, Ashfaq S, Rhodes SD, Weintraub WS, et al. Association between novel oxidative stress markers and C-reactive protein among adults without clinical coronary heart disease. Atherosclerosis. 2005 Jan;178(1):115-21.
69. Davi G, Guagnano MT, Ciabattoni G, Basili S, Falco A, Marinopiccoli M, et al. Platelet activation in obese women: role of inflammation and oxidant stress. JAMA : the journal of the American Medical Association. 2002 Oct 23-30;288(16):2008-14.
70. Pepys MB, Hirschfield GM. C-reactive protein: a critical update. J Clin Invest. 2003 Jun;111(12):1805-12.
71. Lloyd-Jones DM, Liu K, Tian L, Greenland P. Narrative review: Assessment of C-reactive protein in risk prediction for cardiovascular disease. Ann Intern Med. 2006 Jul 4;145(1):35-42.
72. Vermeire S, Van Assche G, Rutgeerts P. The role of C-reactive protein as an inflammatory marker in gastrointestinal diseases. Nat Clin Pract Gastroenterol Hepatol. 2005 Dec;2(12):580-6.
73. Tsilidis KK, Branchini C, Guallar E, Helzlsouer KJ, Erlinger TP, Platz EA. C-reactive protein and colorectal cancer risk: a systematic review of prospective studies. Int J Cancer. 2008 Sep 1;123(5):1133-40.
74. Heikkila K, Harris R, Lowe G, Rumley A, Yarnell J, Gallacher J, et al. Associations of circulating C-reactive protein and interleukin-6 with cancer risk: findings from two prospective cohorts and a meta-analysis. Cancer Causes Control. 2009 Feb;20(1):15-26.
75. Aleksandrova K, Jenab M, Boeing H, Jansen E, Bueno-de-Mesquita HB, Rinaldi S, et al. Circulating C-reactive protein concentrations and risks of colon and rectal cancer: a nested case-control study within the European Prospective Investigation into Cancer and Nutrition. Am J Epidemiol. 2010 Aug 15;172(4):407-18.
76. Tsilidis KK, Erlinger TP, Rifai N, Hoffman S, Hoffman-Bolton J, Helzlsouer KJ, et al. C-reactive protein and colorectal adenoma in the CLUE II cohort. Cancer Causes Control. 2008 Aug;19(6):559-67.

77. Kim S, Keku TO, Martin C, Galanko J, Woosley JT, Schroeder JC, et al. Circulating levels of inflammatory cytokines and risk of colorectal adenomas. Cancer Res. 2008 Jan 1;68(1):323-8.
78. Ognjanovic S, Yamamoto J, Saltzman B, Franke A, Ognjanovic M, Yokochi L, et al. Serum CRP and IL-6, genetic variants and risk of colorectal adenoma in a multiethnic population. Cancer Causes Control. 2010 Jul;21(7):1131-8.
79. Boyapati SM, Bostick RM, McGlynn KA, Fina MF, Roufail WM, Geisinger KR, et al. Calcium, vitamin D, and risk for colorectal adenoma: dependency on vitamin D receptor BsmI polymorphism and nonsteroidal anti-inflammatory drug use? Cancer Epidemiol Biomarkers Prev. 2003 Jul;12(7):631-7.
80. Daniel CR, Bostick RM, Flanders WD, Long Q, Fedirko V, Sidelnikov E, et al. TGF-alpha expression as a potential biomarker of risk within the normal-appearing colorectal mucosa of patients with and without incident sporadic adenoma. Cancer Epidemiol Biomarkers Prev. [Research Support, Non-U.S. Gov't]. 2009 Jan;18(1):65-73.
81. Fedirko V, Bostick RM, Goodman M, Flanders WD, Gross MD. Blood 25-hydroxyvitamin D3 concentrations and incident sporadic colorectal adenoma risk: a pooled case-control study. Am J Epidemiol. 2010 Sep 1;172(5):489-500.
82. Greenland S, Pearl J, Robins J. Causal diagrams for epidemiologic research. Epidemiology. 1999;10:37-48.
83. Muthén LK, Muthén BO. Mplus User's Guide. Seventh ed. Los Angeles, CA: Muthén & Muthén; (1998-2012).
84. Cancer Facts & Figures. Atlanta: American Cancer Society 2012.
85. Haggar FA, Boushey RP. Colorectal cancer epidemiology: incidence, mortality, survival, and risk factors. Clin Colon Rectal Surg. 2009 Nov;22(4):191-7.
86. World Cancer Research Fund/American Institute for Cancer Research. Washington DC: AICR; 2007.
87. Sato T, Takeda H, Otake S, Yokozawa J, Nishise S, Fujishima S, et al. Increased plasma levels of 8-hydroxydeoxyguanosine are associated with development of colorectal tumors. J Clin Biochem Nutr. 2010 Jul;47(1):59-63.
88. Halliwell B, Rafter J, Jenner A. Health promotion by flavonoids, tocopherols, tocotrienols, and other phenols: direct or indirect effects? Antioxidant or not? Am J Clin Nutr. 2005 Jan;81(1 Suppl):268S-76S.
89. Hadziabdic MO, Bozikov V, Pavic E, Romic Z. The antioxidative protecting role of the Mediterranean diet. Coll Antropol. 2012 Dec;36(4):1427-34.
90. Papaioannou D, Cooper KL, Carroll C, Hind D, Squires H, Tappenden P, et al. Antioxidants in the chemoprevention of colorectal cancer and colorectal adenomas in the general population: a systematic review and meta-analysis. Colorectal Dis. 2011 Oct;13(10):1085-99.
91. Jung S, Wu K, Giovannucci E, Spiegelman D, Willett WC, Smith-Warner SA. Carotenoid intake and risk of colorectal adenomas in a cohort of male health professionals. Cancer Causes Control. 2013 Apr;24(4):705-17.
92. Greenberg ER, Baron JA, Tosteson TD, Freeman DH, Jr., Beck GJ, Bond JH, et al. A clinical trial of antioxidant vitamins to prevent colorectal adenoma. Polyp Prevention Study Group. The New England journal of medicine. 1994 Jul 21;331(3):141-7.
93. Poulsen HE, Prieme H, Loft S. Role of oxidative DNA damage in cancer initiation and promotion. Eur J Cancer Prev. 1998 Feb;7(1):9-16.
94. Klaunig JE, Kamendulis LM. The role of oxidative stress in carcinogenesis. Annu Rev Pharmacol Toxicol. 2004;44:239-67.
95. Garland M, Willett WC, Manson JE, Hunter DJ. Antioxidant micronutrients and breast cancer. J Am Coll Nutr. 1993 Aug;12(4):400-11.

96. Block G. The Data Support a Role for Antioxidants in Reducing Cancer Risk. Nutrition Reviews. 1992;50(7):207-13.
97. American Institute for Cancer Research., World Cancer Research Fund. Food, nutrition, physical activity and the prevention of cancer: a global perspective : a project of World Cancer Research Fund International. Washington, D.C.: American Institute for Cancer Research; 2007.
98. Giovannucci E. Tomatoes, Tomato-Based Products, Lycopene, and Cancer: Review of the Epidemiologic Literature. Journal of the National Cancer Institute. 1999;91(4):317-31.
99. Byers T, Perry G. Dietary carotenes, vitamin C, and vitamin E as protective antioxidants in human cancers. Annu Rev Nutr. 1992;12:139-59.
100. van der Vaart H, Postma DS, Timens W, ten Hacken NH. Acute effects of cigarette smoke on inflammation and oxidative stress: a review. Thorax. 2004 Aug;59(8):713-21.
101. Das SK, Vasudevan DM. Alcohol-induced oxidative stress. Life Sciences. 2007;81(3):177-87.
102. Zima T, Fialova L, Mestek O, Janebova M, Crkovska J, Malbohan I, et al. Oxidative stress, metabolism of ethanol and alcohol-related diseases. J Biomed Sci. 2001 Jan-Feb;8(1):59-70.
103. Lugasi A, Hóvári J. Antioxidant properties of commercial alcoholic and nonalcoholic beverages. Food / Nahrung. 2003;47(2):79-86.
104. Arranz S, Chiva-Blanch G, Valderas-Martinez P, Medina-Remon A, Lamuela-Raventos RM, Estruch R. Wine, beer, alcohol and polyphenols on cardiovascular disease and cancer. Nutrients. 2012 Jul;4(7):759-81.
105. Keaney JF, Jr., Larson MG, Vasan RS, Wilson PW, Lipinska I, Corey D, et al. Obesity and systemic oxidative stress: clinical correlates of oxidative stress in the Framingham Study. Arterioscler Thromb Vasc Biol. 2003 Mar 1;23(3):434-9.
106. Finkel T, Holbrook NJ. Oxidants, oxidative stress and the biology of ageing. Nature. 2000 Nov 9;408(6809):239-47.
107. Vincent HK, Taylor AG. Biomarkers and potential mechanisms of obesity-induced oxidant stress in humans. Int J Obes (Lond). 2006 Mar;30(3):400-18.
108. Podhaisky H-P, Abate A, Polte T, Oberle S, Schröder H. Aspirin protects endothelial cells from oxidative stress – possible synergism with vitamin E. FEBS Letters. 1997;417(3):349-51.
109. Stahl W, Sies H. Antioxidant activity of carotenoids. Mol Aspects Med. 2003 Dec;24(6):345-51.
110. Halliwell B. Effect of diet on cancer development: is oxidative DNA damage a biomarker? Free Radic Biol Med. 2002 May 15;32(10):968-74.
111. Maiese K. Marking the onset of oxidative stress: biomarkers and novel strategies. Oxid Med Cell Longev. 2009 Jan-Mar;2(1):1.
112. Dalle-Donne I, Rossi R, Colombo R, Giustarini D, Milzani A. Biomarkers of Oxidative Damage in Human Disease. Clinical Chemistry. 2006;52(4):601-23.
113. Tee ES, Jr., Khor S. Simultaneous determination of retinol, carotenoids and tocopherols in human serum by high pressure liquid chromatography. Malays J Nutr. 1995 Sep;1(2):151-70.
114. Morrow JD, Roberts LJ, 2nd. Mass spectrometric quantification of F2-isoprostanes in biological fluids and tissues as measure of oxidant stress. Methods Enzymol. 1999;300:3-12.
115. Shimasaki H. Assay of fluorescent lipid peroxidation products. Methods Enzymol. 1994;233:338-46.
116. Meredith W. Measurement Invariance, factor analysis and factorial invariance. Psychometrika. 1993;58(4):525-43.
117. Millsap RE. Statistical approaches to measurement invariance. New York: Psychology Press; 2011.

118. Meredith W, Teresi JA. An essay on measurement and factorial invariance. Med Care. 2006 Nov;44(11 Suppl 3):S69-77.
119. Byrne BM, Shavelson RJ, Muthén B. Testing for the Equivalence of Factor Covariance and Mean Structure: The Issue of Partial Measurement Invariance. Psychological Bullentin. 1989;105(3):456-66.
120. Halliwell B, Lee CY. Using isoprostanes as biomarkers of oxidative stress: some rarely considered issues. Antioxid Redox Signal. 2010 Jul 15;13(2):145-56.
121. Scholz H, Yndestad A, Damas JK, Waehre T, Tonstad S, Aukrust P, et al. 8-isoprostane increases expression of interleukin-8 in human macrophages through activation of mitogen-activated protein kinases. Cardiovasc Res. 2003 Oct 1;59(4):945-54.
122. Clay Montier LL, Deng JJ, Bai Y. Number matters: control of mammalian mitochondrial DNA copy number. J Genet Genomics. 2009 Mar;36(3):125-31.
123. Poyton RO, Ball KA, Castello PR. Mitochondrial generation of free radicals and hypoxic signaling. Trends Endocrinol Metab. 2009 Sep;20(7):332-40.
124. de Ferranti S, Rifai N. C-reactive protein and cardiovascular disease: a review of risk prediction and interventions. Clin Chim Acta. 2002 Mar;317(1-2):1-15.
125. Kregel KC, Zhang HJ. An integrated view of oxidative stress in aging: basic mechanisms, functional effects, and pathological considerations. Am J Physiol Regul Integr Comp Physiol. 2007 Jan;292(1):R18-36.
126. Nielsen F, Mikkelsen BB, Nielsen JB, Andersen HR, Grandjean P. Plasma malondialdehyde as biomarker for oxidative stress: reference interval and effects of life-style factors. Clin Chem. 1997 Jul;43(7):1209-14.
127. Polidori MC, Praticó D, Ingegni T, Mariani E, Spazzafumo L, Del Sindaco P, et al. Effects of vitamin C and aspirin in ischemic stroke-related lipid peroxidation: Results of the AVASAS (Aspirin Versus Ascorbic acid plus Aspirin in Stroke) Study. BioFactors. 2005;24(1):265-74.
128. Urquiaga I, Strobel P, Perez D, Martinez C, Cuevas A, Castillo O, et al. Mediterranean diet and red wine protect against oxidative damage in young volunteers. Atherosclerosis. 2010 Aug;211(2):694-9.
129. Sies H, Stahl W. Vitamins E and C, beta-carotene, and other carotenoids as antioxidants. Am J Clin Nutr. 1995 Dec;62(6 Suppl):1315S-21S.
130. Mayne ST. Antioxidant nutrients and chronic disease: use of biomarkers of exposure and oxidative stress status in epidemiologic research. J Nutr. 2003 Mar;133 Suppl 3:933S-40S.
131. Basu A, Imrhan V. Tomatoes versus lycopene in oxidative stress and carcinogenesis: conclusions from clinical trials. Eur J Clin Nutr. 2007 Mar;61(3):295-303.
132. Miller AA, De Silva TM, Jackman KA, Sobey CG. Effect of gender and sex hormones on vascular oxidative stress. Clin Exp Pharmacol Physiol. 2007 Oct;34(10):1037-43.
133. Tiidus PM. Estrogen and gender effects on muscle damage, inflammation, and oxidative stress. Can J Appl Physiol. 2000 Aug;25(4):274-87.
134. Helmersson J, Mattsson P, Basu S. Prostaglandin F(2alpha) metabolite and F(2)-isoprostane excretion rates in migraine. Clin Sci (Lond). 2002 Jan;102(1):39-43.
135. Rifai N, Ridker PM. Population distributions of C-reactive protein in apparently healthy men and women in the United States: implication for clinical interpretation. Clin Chem. 2003 Apr;49(4):666-9.
136. Khera A, McGuire DK, Murphy SA, Stanek HG, Das SR, Vongpatanasin W, et al. Race and gender differences in C-reactive protein levels. J Am Coll Cardiol. 2005 Aug 2;46(3):464-9.
137. McConnell JP, Branum EL, Ballman KV, Lagerstedt SA, Katzmann JA, Jaffe AS. Gender differences in C-reactive protein concentrations-confirmation with two sensitive methods. Clin Chem Lab Med. 2002 Jan;40(1):56-9.

138. Oberley TD. Oxidative damage and cancer. Am J Pathol. 2002 Feb;160(2):403-8.
139. Valko M, Leibfritz D, Moncol J, Cronin MT, Mazur M, Telser J. Free radicals and antioxidants in normal physiological functions and human disease. Int J Biochem Cell Biol. 2007;39(1):44-84.
140. Halliwell B, Grootveld M. The measurement of free radical reactions in humans: Some thoughts for future experimentation. FEBS Letters. 1987;213(1):9-14.
141. Veglia F, Cavalca V, Tremoli E. OXY-SCORE: a global index to improve evaluation of oxidative stress by combining pro- and antioxidant markers. Methods Mol Biol. 2010;594:197-213.
142. Dash C, Goodman M, Flanders WD, Mink PJ, McCullough ML, Bostick RM. Using pathway-specific comprehensive exposure scores in epidemiology: application to oxidative balance in a pooled case-control study of incident, sporadic colorectal adenomas. Am J Epidemiol. 2013 Aug 15;178(4):610-24.
143. Eldridge RC, Flanders WD, Bostick R, Fedirko V, Goodman M. Using multiple determinants and biomarkers to obtain a better measurement of oxidative stress: a latent variable structural equation model approach. 2014.
144. Willett WC, Sampson L, Stampfer MJ, Rosner B, Bain C, Witschi J, et al. Reproducibility and validity of a semiquantitative food frequency questionnaire. Am J Epidemiol. 1985 Jul;122(1):51-65.
145. Willett WC, Sampson L, Browne ML, Stampfer MJ, Rosner B, Hennekens CH, et al. The use of a self-administered questionnaire to assess diet four years in the past. Am J Epidemiol. 1988 Jan;127(1):188-99.
146. SAS, Institute, Inc. Base SAS 9.3 Procedures Guide. Cary, NC: SAS Institute Inc.; 2011.
147. Muthen LK, Muthen BO. Mplus User's Guide 7th edition. Los Angeles, CA: Muthen & Muthen; (1998-2012).
148. Vanderweele TJ, Vansteelandt S. Odds ratios for mediation analysis for a dichotomous outcome. Am J Epidemiol. 2010 Dec 15;172(12):1339-48.
149. Cole BF, Logan RF, Halabi S, Benamouzig R, Sandler RS, Grainge MJ, et al. Aspirin for the chemoprevention of colorectal adenomas: meta-analysis of the randomized trials. J Natl Cancer Inst. 2009 Feb 18;101(4):256-66.
150. Giovannucci E, Ascherio A, Rimm EB, Colditz GA, Stampfer MJ, Willett WC. Physical Activity, Obesity, and Risk for Colon Cancer and Adenoma in Men. Annals of Internal Medicine. 1995;122(5):327-34.
151. Giovannucci E, Colditz G, Stampfer M, Willett W. Physical activity, obesity, and risk of colorectal adenoma in women (United States). Cancer Causes & Control. 1996 1996/03/01;7(2):253-63.
152. Ridker PM, Hennekens CH, Buring JE, Rifai N. C-reactive protein and other markers of inflammation in the prediction of cardiovascular disease in women. The New England journal of medicine. 2000 Mar 23;342(12):836-43.
153. Hafeman DM, VanderWeele TJ. Alternative assumptions for the identification of direct and indirect effects. Epidemiology. 2011;22(6):753-64.
154. MacKinnon DP, Fairchild AJ, Fritz MS. Mediation analysis. Annu Rev Psychol. 2007;58:593-614.
155. VanderWeele TJ. Invited commentary: structural equation models and epidemiologic analysis. Am J Epidemiol. 2012 Oct 1;176(7):608-12.
156. Maffei F, Angelini S, Forti GC, Hrelia P. Blood Biomarkers Linked to Oxidative Stress and Chronic Inflammation for Risk Assessment of Colorectal Neoplasia. Curr Colorectal Cancer Rep. 2013;9(1):85-94.

157. Rothman K, Greenland S, Lash T. Modern Epidemiology. 3rd ed. Philadelphia: Lippincott Williams & Wilkins; 2008.
158. Hernan MA, Hernandez-Diaz S, Robins JM. A structural approach to selection bias. Epidemiology. 2004 2004 Sep;15(5):615-25.
159. Kopple JD. The phenomenon of altered risk factor patterns or reverse epidemiology in persons with advanced chronic kidney failure. Am J Clin Nutr. 2005 Jun;81(6):1257-66.
160. Levin NW, Handelman GJ, Coresh J, Port FK, Kaysen GA. Reverse epidemiology: a confusing, confounding, and inaccurate term. Semin Dial. 2007 Nov-Dec;20(6):586-92.
161. Preston SH, Stokes A. Obesity paradox: conditioning on disease enhances biases in estimating the mortality risks of obesity. Epidemiology. 2014 May;25(3):454-61.
162. Flanders WD, Eldridge RC, McClellan W. A Nearly Unavoidable Mechanism for Collider Bias with Index-Event Studies. Epidemiology. 2014;In Press.
163. Flanders WD, Eldridge RC. Summary of Relationships between Exchangeability, Biasing Paths and Bias. European Journal of Epidemiology. 2014;In Press.
164. Tu Y-K. Commentary: Is structural equation modelling a step forward for epidemiologists? International Journal of Epidemiology. 2009;38(2):549-51.
165. Baltar VT, Xun WW, Johansson M, Ferrari P, Chuang SC, Relton C, et al. A structural equation modelling approach to explore the role of B vitamins and immune markers in lung cancer risk. Eur J Epidemiol. 2013 Aug;28(8):677-88.