Challenges Associated with Statistical Analysis in the Presence of Sparse Data and Applications to Alternative Tobacco Product Research

Daniel Szydlo

A thesis submitted in partial fulfillment of the requirements of the degree of

Master of Science

University of Washington

2013

Committee:

Susanne May

Gerald van Belle

Program Authorized to Offer Degree:

School of Public Health - Biostatistics

University of Washington

Abstract

Challenges Associated with Statistical Analysis in the Presence of Sparse Data and Applications to Alternative Tobacco Product Research

Daniel Szydlo

Chair of the Supervisory Committee:

Susanne May

Biostatistics

Rarely observed covariate combinations, or "sparsity" is a phenomenon associated with research concerning the health risks of alternative-use (non-combusted tobacco products (AUPs)). Of particular concern is sparsity relating to AUP users who do not currently or formerly use other tobacco products. This thesis aims to identify reasons why sparsity is a concern, the effect that sparsity can have on statistical inference, and potential appropriate approaches in the presence of sparsity. Special attention will be paid to scenarios in which sparsity can lead to inference that results in estimates of the AUP effect that are in the opposite direction of the true effect (e.g. found to be harmful when truly beneficial) and to be in an opposite direction related to the cigarette effect (e.g. found to be less harmful than cigarettes when truly more harmful). The impact of sparsity will be assessed primarily by constructing examples from both case-control and cohort studies and investigating the results from common statistical modeling methods under sparse and non-sparse conditions. These examples will include hypothetical examples constructed to approximate real world study design as well as data from a published study of an AUP. These examples will focus on issues of sparsity in relation to interaction assumptions and model scale assumptions. Conditional parameter estimates can vary widely from the marginal estimates for that parameter. Data sets with few subjects who use AUPs without also using cigarettes have reduced power to detect interaction. When scale or interaction assumptions are violated estimation of incidence rate or parameter values can be biased. This bias can be such that conclusions from analysis of sparse data sets can be misleading. These issues can cause AUP use to be estimated as beneficial when it is in truth harmful, or as less harmful than cigarettes when in truth it is more harmful. These issues are of such severity that we, if it is not possible to oversample the sparse categories, recommend restricting analysis to subgroups in which sparsity is unlikely to be a concern.

Contents

List of Figures

List of Tables

1 Introduction

With the Family Smoking Prevention and Tobacco Control Act (FSPTCA), signed into law in 2009, assigning responsibility for regulation of tobacco products to the Food and Drug Administration (FDA) there is a need for public health research that investigates the health risks associated with alternative-use (non-combusted) tobacco products (AUPs) [1]. AUP research is potentially complicated by certain characteristics encountered in research into the health effects of tobacco use. These characteristics include concomitant use of an AUP along with combusted cigarettes, rarity of initiation of AUP use without a prior history of cigarette use, and difficulties in summarizing longitudinal tobacco use patterns and their effect on health outcomes. These concerns are especially relevant since it unethical to randomize subjects to use tobacco, so research into the health effects of tobacco will primarily be observational, case-control, or cohort designs.

Many studies have found evidence of concomitant use patterns between cigarettes and AUPs, or a tendency for AUP users to have had a history of cigarette use [3] [5] [10] [17] [21] [22]. The Center for Disease Control and Prevention (CDC) National Health Interview Surveys (NHIS) from 1991 and 2006 estimated that 2.9% of adults in the United States regularly used smokeless tobacco, however, this percentage was approximately 2.2% among non-smokers compared to 3.9% among those with a history of smoking [3] [5]. A study of United States Air Force (USAF) recruits showed a similar trend, with current use of smokeless tobacco reported as 9.6% among former smokers compared to 4.6% and 4.1% among current and never smokers, respectively [10]. A study of middle and high school students in Massachusetts found that 84.8% of current users of alternative tobacco products also had some history of cigarette use and 81.4% of lifetime AUP users had some history of cigarette use [17]. Researchers investigating the association between tobacco use and renal cell carcinoma in residents of Los Angeles noted that "most individuals who used noncigarette tobacco products had also smoked cigarettes regularly" [22]. A study of concomitant use of cigarettes and smokeless tobacco used data from a variety of sources to estimate that 0.6-0.8% of the U.S. adult population used both smokeless tobacco and cigarettes, and that at least 25% of current smokeless tobacco users also currently used cigarettes [21].

These findings demonstrate a fairly intuitive relationship between use of cigarettes and AUPs, with lifetime users of cigarettes being more likely to also have some lifetime use of an AUP than non-smokers. As such, a study investigating the health risks associated with AUP use may have few subjects in the category of AUP users with no history of cigarette use. We refer to this condition as "sparsity". If cigarette use and AUP use are simply categorized as ever users vs. never users, we are concerned that we may observe sparsity for the cell corresponding to AUP users who did not smoke cigarettes, as shown in Table 1.

Table 1: Sparsity in AUP Status by Cigarette Status

	Cigarettes?	
AUPs?	No	Yes
No	Not sparse	Not sparse
Yes	**Sparse**	Not sparse

More specifically, based on the studies cited above we believe that it is reasonable to expect something along the lines of what is shown in Table 2 [4].

Table 2: Reasonable Expected Percentages of AUP Use and Cigarette Use

	Cigarettes?	
AUPs?	No	Yes
No	1-4%	35-45%
Yes	**0.2-1%**	50-60%

We believe that two of the major questions of interest in AUP research are whether or not AUP use is hazardous, and whether AUP use is less hazardous when compared to cigarette use. Both of these questions imply an interest in the effects of AUP use by itself, rather than AUP use accompanied by a reduction in cigarette smoking or among former smokers. We believe that if one is interested in assessing the health impact of AUP use in the absence of current or historical cigarette use, then observing few subjects who used only AUPs will affect the ability to extend inference to this category. This category of non-smoking AUP users can be thought of as representing the main effect of an AUP on a health outcome. A study of this category of users will provide information on the association between AUP use and outcome among those who did not smoke cigarettes.

Recommendations regarding how to proceed if there were no data at all for a covariate combination (a condition referred to variously as "missing cells" or "empty cells"; we will use "empty cells" in order to avoid confusion with "missing data") have been previously established [6] [11]. In a discussion of analyses for a two-way treatment structure in which there is no data for one of the treatment combinations, Milliken and Johnson state that "certain hypotheses cannot be tested without making some additional assumptions about the parameters in the model. Hypotheses involving parameters corresponding to the missing cells generally cannot be tested" [11]. To illustrate this point they provide a hypothetical example of a two-way experiment with two three-level treatments T and B, but with no subjects receiving the combinations of T_1B_2 or T_3B_3, as shown in Table 3 [11].

Table 3: Two-Way Experiment with Missing Treatment Combinations from Milliken [11]

	B_1	B_2	B_3
T_1	X		X
T_2	X	X	X
T_3	X	X	

In this example it is not possible to make inference not only on the effect of those specific combinations, but also on the main effects for T and B unless we make untestable interaction/effect modification (which we will refer to as "interaction") assumptions regarding their relationship [11]. This applies not only to a hypothesis regarding the effect of T_1B_2 but also to hypotheses about the marginal effects of T_1, T_3, B_2, and B_3, since those marginal effects are calculated over the empty cells. If hypotheses regarding these margins are tested, the validity of those tests will depend upon untestable assumptions regarding the presence or absence of interaction between those treatments. Depending on the scenario it may be entirely reasonable to assume no interaction and proceed with an estimation of the main effects for the two treatments, but this will not always be the case. "All too often we find experimenters willing to assume no interaction exists between the treatment levels in any of their experiments mainly because they do not understand how to deal with such

an interaction or because they believe they are not interested in it" [11].

What is not as well established is what, if any, analyses are appropriate when covariate combinations are not entirely empty, but rather are sparsely populated. Of particular interest is the scenario where data is sparse for the exposure of interest in the absence of a second related exposure, or, in terms of our focus, when AUP use is very rarely observed without a history of cigarette use. In attempting to address this question we are also interested in what impact sparsity has on statistical inference in cases where it is ignored or not noticed. We hypothesize that analysis in the presence of sparse data relies on strong and potentially untestable assumptions. Specifically, we hypothesize that, since modeling of the main effect for AUP use depends on having data for that main effect, analysis when AUP use without a history of cigarette use data is sparse will be prone to bias and susceptible to violations of scale, especially when compared to non-sparse data sets. We believe that there are substantial limitations to the estimation of the main effect for AUP use when standard analysis tools (e.g., logistic regression) are used and that use of these methods can lead to incorrect results and interpretation.

As long as all cells contain some data it is possible to estimate all main and interaction effects and standard methods will give unbiased estimators provided all assumptions are met [16]. However, while these estimates will be unbiased, sparse data will lead to wide confidence interval estimates for cells containing sparse data due to small n for these cells. This also does not consider that assumptions on which asymptotic unbiasedness are dependent may be untestable. We hypothesize that it will be difficult or impossible to verify that assumptions involving the sparse data cells have been met, and that if they have been violated the main and interaction effect estimates can no longer be assumed to be unbiased. Identification of when data is sparse, limitations to inference that occur as a result of sparsity, and guidelines for conducting statistical analyses (or when statistical analyses are not useful) in the presence of sparsity will be addressed in this thesis.

We are concerned that the effects of sparsity will be magnified when observed in conjunction with other complicating factors encountered in tobacco research. A person's tobacco use can be represented many ways in a regression model, with common approaches including dichotomous (ever used vs. never used, ≤ 20 years duration vs. > 20, $\leq$20 cigarettes/day vs. > 20, ≤ 20 pack-years vs. $>$20), trichotomous (never used vs. current user vs. former user, trichotomous categorizations of duration, number per day, or pack years), or finer categorizations (no tobacco use vs. only cigarettes vs cigarettes and other vs only other) [8] [15] [23]. In published tobacco research that we examined, the highest order categorization scheme used five different values: never smoker vs. former vs. relapsed former vs. abstinent vs. persistent [7]. These categorizations can be used to simplify the data for analysis, because they are thought to better represent the relationship between exposure and outcome, or because a small number of categories will be easier to interpret from a public health perspective [18]. However, the categorizations and cutpoints sometimes involve arbitrary decisions and we believe they have the potential to fail to accurately reflect the true underlying mechanism by which tobacco is related to a health outcome. We hypothesize that overly coarse or outright misclassification can lead to biased regression estimates and that this impact can be magnified by the presence of sparsity.

As alluded to previously, it is important to note that sparsity and empty cells are different than missing data. In the case of the combination of AUP and cigarette use, missing data would occur if there were subjects in the data set whose status for these tobacco use variables was unknown, whereas sparse/empty data would occur if there were few subjects whose variables values fit into that category. This means that recommendations regarding data missing not at random or in-

formative missingness will not be applicable to situations involving sparse and empty data. It is also important to note that we are focused on sparsity and emptiness for combinations of multiple covariates, rather than for a single covariate. We are not concerned with whether AUP use is expected to be very rare in the population, but rather that it is very rare among persons who do not currently use or have a history of using cigarettes, as we feel that there is a better understanding of the implications of univariate sparsity. We also will not deal with sparsity involving combinations of more than two covariates, as such higher level interactions are rarely observed among in research that does not focus on genetic factors, and since issues related to two factors will generalize to more than two factors [18].

We will address these concerns primarily through the use of empirical examples. These will be based on research studies that we have encountered in the literature and on future studies where we hope to provide guidance regarding study design and questions of interest. We focused on two studies: an existing study that investigated the health risks associated with cigarette and AUP use in Sweden, and the future study that was commissioned as a response to the FDA mandate.

Researchers in Sweden conducted a case-control study with oropharyngeal squamous cell carcinoma (OOSCC) the health outcome of interest [14]. Data were collected on subjects' use of Swedish moist snuff and cigarettes [14]. 132 cases and 320 controls were enrolled from southern Sweden [14]. Average daily consumption of cigarettes was gathered and total lifetime consumption was calculated by multiplying average consumption by duration of use [14]. For the analysis, these continuous variables were categorized by both average (never smoked vs. 1-10 cigarettes/day vs. 11-20 vs. $>$ 20) and total (never smoked vs. $<$125 kilograms (kg) during lifetime of individual vs. 125-250 kg vs. $>$250 kg) consumption [14]. Moist snuff use was was dichotomized (never used vs. ever used) [14]. Some other categorizations were investigated as well, but we will not focus on them in our analyses [14]. Multivariable analysis of average cigarette consumption (adjusted for alcohol use) estimated odds ratios of 1.1 (95% CI: 0.6 - 2.1) for smoking 1-10 cigarettes/day, 2.4 (1.3 - 4.1) for 11-20 cigarettes/day, and 2.8 (1.3 - 6.1) for greater than 20 cigarettes per day [14]. Univariate analysis of moist snuff usage was not statistically significant (OR = 0.7, 0.4 - 1.2) nor was snuff usage significant in a multivariate analysis (OR = 0.7, 0.3 - 1.3) adjusting for alcohol and cigarette consumption [14]. The authors did not discuss the possibility of interaction existing between cigarettes and moist snuff, nor was data presented for the observed frequencies of moist snuff use status by cigarette consumption. We will investigate the implications of this lack of consideration of interaction, hypothesizing that it is possible that interaction could have been observed and that it would have been possible to have a data set which was highly suggestive of an interaction between cigarettes and moist snuff and which still resulted in the marginal odds ratios presented in the paper.

As mentioned, with the FSPTCA giving the FDA mandate on the regulation of tobacco products, there will need to be studies that investigate the health effects of those products. There could also be increased emphasis on identifying alternatives to cigarette use such as AUPs. As a result we believe that it is vitally important to better understand the impact of forms of sparsity that may be observed in these studies. For instance, as a response to the FDA mandate the National Institute of Health (NIH) issued a Request for Proposal (RFP) in January 2011, seeking proposals for a cohort study focusing on AUP research [13]. This RFP sought to fund a study that would initially focus on "near-term behavioral and health effects" related to regulatory changes stemming from the FDA mandate, with the expectation that over time the data set would allow for longitudinal analyses of long-term health risks [13]. The Population Assessment of Tobacco and Health (PATH) study was awarded funding [13]. This study, run by Westat, is a longitudinal cohort study that will monitor

tobacco usage patterns in more than 40,000 smokers and non-smokers, focusing on determining the influence of FDA regulation on tobacco products as well as the impact of the mandate on long-term tobacco-related morbidity and mortality [12] [20]. As this study will not begin until Fall 2013 there is no current data from this study [12]. Investigators performing such studies such as this will need to be aware of ways in which sparsity can affect the ability to perform inference.

2 Description of Examples

Our approach to investigating the effects of sparsity will involve constructing examples that attempt to replicate scenarios that approximate potential real world tobacco studies, and assess the impact of sparsity and other concerns in these scenarios. We will use a variety of examples in order to investigate the impact of sparsity. These will include some purely hypothetical examples as well as examples that are based on existing studies. Both case-control and cohort studies will be examined. In this section we will describe how these examples will be designed, what statistical methods will be employed, and what these examples are intended to help us investigate and illustrate.

Example 1: Relationship Between Conditional and Marginal Odds Ratio Estimates in the Presence of Interaction
The purpose of our first example is to illustrate the relationship (or lack thereof) between marginal and conditional odds ratios. We are specifically interested in what differences can exist between estimates for the effect of a covariate depending on whether interaction is assumed to be present. Data will be generated from a hypothetical case-control study estimating the association between tobacco use and esophageal cancer. The study will consist of 180 cases and 360 controls. The covariates we will consider are AUP use and cigarette use, both treated as dichotomous variables (ever use vs. never use). Subjects' esophageal cancer will be dependent on these tobacco statuses and no other covariate.

We will generate data from six different scenarios under the conditions described above, all with distributions of observed frequencies of combinations of the three collected variables such that the conditional odds ratio estimating the association between AUP use and esophageal cancer among cigarette abstainers is 1.2. Specifically, the number of cases who used AUPs but not cigarettes, the number of cases who did not use AUPs and did not use cigarettes, the number of controls who used AUPs and did not use cigarettes, and the number of controls who did not use AUPs and did not use cigarettes will all be fixed (45, 45, 100, and 120, respectively). In order to reduce the number of unknowns we will also fix the number of cases who used cigarettes to be 45 for both AUP users and non-users. This leaves 140 non-smoking controls to be distributed between AUP users and non-users, which are shown as cells a and b in Table 4. This also results in a fixed estimate of the unadjusted marginal odds ratio associated with cigarette use for all six scenarios.

Table 4: Example 1, Case Status by Cigarette and AUP Use

	Cases		Controls	
	AUP		AUP	
Cigarette status	Yes	No	Yes	No
Yes	45	45	a	b
No	45	45	100	120

The six scenarios will be generated by varying the observed values for a and b. These cell counts will be selected in order to attempt to satisfy the following scenarios: marginal AUP effect is greater than both the effect of AUP use among non-smokers as well as the effect of cigarette use (Scenario 1), marginal AUP effect is between the effect of AUP use among non-smokers and the effect of cigarette use (Scenario 2), marginal AUP effect is equal to the effect of AUP use among non-smokers (Scenario 3), marginal effect of AUP use is less than the effect of AUP use among non-smokers but greater than no effect (Scenario 4), no marginal effect of AUP use (Scenario 5),

and a marginal effect of AUP use than is less than no effect, or is beneficial (Scenario 6). We will calculate the unadjusted odds ratio for marginal AUP use in each of these data sets. These will be compared to the odds ratio for AUP use among non-smokers, which, as noted, we have fixed to be 1.2.

The number of cases and controls, as well as the case-control study design itself, were motivated by a study of the association between moist snuff use, tobacco use, and OOSCC that used a roughly similar study design [14]. We intended to mimic the type of study that could be implemented in actual research, but the implications of these examples should be generalizable beyond this specific study setting. In Example 2 we will examine the moist snuff study more closely and investigate some potential flaws related to interaction assumptions in the study analysis.

Example 2: Implications of Violated Interaction Assumptions, As Seen in a Case-Control Study

We will base Example 2 on a case-control study that investigated the association between tobacco use (cigarettes and Swedish moist snuff) and OOSCC but that did not consider the possibility of interaction between AUP use and cigarette use [14]. Background and results of this study were described in the Introduction. We aim to determine whether significant evidence of interaction could have existed in this study, as well as the effect sizes of parameter estimates made when assuming interaction existed. This is similar to the aims of Example 1, but approaches the issue from a different direction; in Example 1 we are investigating the possibility for variation in a marginal odds ratio estimate in a scenario in which the conditional odds ratio estimate for that parameter is fixed and known but in Example 2 we are reversing this and examining the potential for variation in the conditional odds ratio estimates for a variable whose marginal odds ratio is known. This example is also directly using data from an actual study, rather than being completely hypothetical.

For our purposes we will focus on the analysis of the association between cigarette use and OOSCC and of the association between moist snuff use and OOSCC, adjusted for cigarette use. From the results presented by the authors we know the cell counts for case-control status by cigarette use status and case-control status by snuff use status, and these are presented in Tables 5 and 6 [14]. We collapsed cigarette use from the four-level categorical definition (never smoked vs. 1-10 cigarettes/day vs. 11-20 cigarettes/day vs. $>$ 20 cigarettes/day) in the paper into a dichotomous variable (never smoked vs. smoked) in order to reduce the number of combinations that were needed to be examined as well as to simplify the interaction between the two forms of tobacco use. Observed frequencies of cigarette use by snuff use, and vice versa, were not included in the published paper and the regression analyses assumed that there was no interaction between cigarette use and snuff use. Since the data presented in the paper did not include a 2x2x2 breakdown of case and control status by both cigarette and moist snuff status we cannot check whether there was evidence of interaction. Instead, we can create hypothetical data sets to represent what potential 2x2x2 distributions could have occurred, calculating the odds ratios for the main effect of moist snuff use and comparing it to the marginal snuff use odds ratio.

We will examine all possible combinations of cigarette, moist snuff, and case statuses that are compatible with the frequencies shown in Tables 5 and 6, applying a logistic regression model that includes moist snuff use (ever vs. never), cigarette use (ever vs. never), and interaction between moist snuff use and cigarette use. This method will allow us to determine all possible conditional odds ratios that could have been observed in the original data set. We will also use the likelihood ratio test (LRT) to determine whether or not each generated data set would have displayed statistically significant evidence of interaction between moist snuff and cigarettes at the

$\alpha = 0.05$ level.

Table 5: Example 2, OOSCC Status and Cigarette Use

Cigarette use	Cases	Controls	Total
Had used	90	166	256
Never used	41	154	195
Total	131 [a]	320	451

[a] 1 case missing cigarette use status

Table 6: Example 2, OOSCC Status and Moist Snuff Use

Moist snuff use	Cases	Controls	Total
Had used	20	65	85
Never used	112	255	367
Total	132	320	452

Example 3: Exploration of Sparsity in Cross-Sectional Cohort Studies
In Examples 1 and 2 we examined the relationship between marginal and conditional parameter estimates in case-control studies, and while some of the scenarios in each example could have involved sparsity, it was not our primary focus. In Example 3 we will focus more explicitly on the presence of sparsity and its effect on inference. It will, however, be important to recall the results of the first two examples during the remainder of this thesis.

In Example 3 we will generate data sets based on what may be observed in the PATH study in order to investigate and illustrate the impact of sparsity. We will construct data sets consisting of 40,000 subjects, collecting the following covariates: cigarette use (no smoking vs. moderate level of smoking vs. heavy level of smoking), AUP use (ever used vs. never used), and, as a later extension, age (treated as continuous taking on 10 possible values). The outcome for these analyses will be referred to as chronic bronchitis, but this is meant to represent a relatively prevalent health outcome, and so rates of bronchitis may be unrealistic. This setup is meant to roughly approximate the basic structure of data that may result from the PATH study.

The data sets will be set such that the observed incidence rate of chronic bronchitis is 0.150 among tobacco abstainers, 0.225 among moderate smokers who did not use the AUP, and 0.300 among heavy smokers who did not use the AUP. We will then determine true observed incidence rates among AUP use categories such that they show no evidence of interaction on a variety of scales (additive scale, multiplicative scale according to odds ratio, multiplicative scale according to relative risk) and such that AUP use has a variety of effects on risk of bronchitis (including beneficial, neutral, or harmful associations). All data sets will be generated for both sparse and non-sparse AUP/cigarette combinations, as defined in Table 7.

Table 7: Example 3, AUP Status by Cigarette Status for Sparse and Non-sparse Data Sets

	Sparse		Non-sparse	
	AUP		AUP	
Cigarettes	Yes	No	Yes	No
Heavy	800 (2.0%)	3300 (8.3%)	6667 (16.7%)	6667 (16.7%)
Moderate	800 (2.0%)	3300 (8.3%)	6667 (16.7%)	6667 (16.7%)
None	200 (0.5%)	31600 (79.0%)	6666 (16.7%)	6666 (16.7%)

Statistical models will be fit to these data sets using linear regression (representing the modeling assumption of an additive scale of association), logistic regression (assuming multiplicative scale of association according to the odds ratio), relative risk regression (assuming multiplicative scale according to the relative risk), and Poisson regression (an approximation of the binomial distribution). For each of these four scale assumptions we will first fit a model that includes interaction between cigarette status and AUP status along with main effects for both types of tobacco use. The interaction terms will be tested for significance at the $\alpha = 0.05$ level using the LRT, and if that test does not meet the criteria for statistical significance a second model will be fitted including only the main effects. We believe that this method is similar to the method that many researchers would employ in a real world analysis of these data, and as such this approach will allow us to observe how a standard statistical analysis could be affected by the presence of sparsity. These models will be referred to as Model A through Model F, and the probabilities (P) of developing chronic bronchitis (CB) are described below. Models A, C, and E assume interaction between cigarette status and AUP status, Models B, D, and F assume no interaction. We note that since relative risk regression and Poisson regression are both generalized linear models that use the log link function, the probabilities of bronchitis for both are represented in Models E and F.

$$
\begin{aligned}
X_1 &= \begin{cases} 1 \text{ if moderate cigarette user} \\ 0 \text{ otherwise} \end{cases} \\
X_2 &= \begin{cases} 1 \text{ if heavy cigarette user} \\ 0 \text{ otherwise} \end{cases} \\
X_3 &= \begin{cases} 1 \text{ if AUP user} \\ 0 \text{ otherwise} \end{cases}
\end{aligned}
$$

Model A: $P(\text{CB}|X_1 = x_1, X_2 = x_2, X_3 = x_3) = \gamma_0 + \gamma_1 x_1 + \gamma_2 x_2 + \gamma_3 x_3 + \gamma_4 x_1 x_3 + \gamma_5 x_2 x_3$

Model B: $P = \gamma_0^* + \gamma_1^* x_1 + \gamma_2^* x_2 + \gamma_3^* x_3$

Model C: $\text{logit}(P) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_1 x_3 + \beta_5 x_2 x_3$

Model D: $\text{logit}(P) = \beta_0^* + \beta_1^* x_1 + \beta_2^* x_2 + \beta_3^* x_3$

Model E: $\log(P) = \delta_0 + \delta_1 x_1 + \delta_2 x_2 + \delta_3 x_3 + \delta_4 x_1 x_3 + \delta_5 x_2 x_3$

Model F: $\log(P) = \delta_0^* + \delta_1^* x_1 + \delta_2^* x_2 + \delta_3^* x_3$

From each model we can determine the estimated incidence rate of bronchitis for each combination of AUP and smoking status by setting x_1, x_2, and x_3 to the desired values, entering the

parameter estimates from that regression model, and solving for P. As an illustration we have provided the equations for estimating incidence rates for the same cell (moderate-smoking AUP-users) in all 6 regression model forms:

$$\begin{aligned}
\text{A: } \widehat{IR}|(x_1 = 1, x_2 = 0, x_3 = 1) &= \gamma_0 + \gamma_1 + \gamma_3 + \gamma_4 \\
\text{B: } \widehat{IR}|(x_1 = 1, x_2 = 0, x_3 = 1) &= \gamma_0^* + \gamma_1^* + \gamma_3^* \\
\text{C: } \widehat{IR}|(x_1 = 1, x_2 = 0, x_3 = 1) &= \text{expit}(\beta_0 + \beta_1 + \beta_3 + \beta_4) \\
\text{D: } \widehat{IR}|(x_1 = 1, x_2 = 0, x_3 = 1) &= \text{expit}(\beta_0^* + \beta_1^* + \beta_3^*) \\
\text{E: } \widehat{IR}|(x_1 = 1, x_2 = 0, x_3 = 1) &= \exp(\delta_0 + \delta_1 + \delta_3 + \delta_4) \\
\text{F: } \widehat{IR}|(x_1 = 1, x_2 = 0, x_3 = 1) &= \exp(\delta_0^* + \delta_1^* + \delta_3^*)
\end{aligned}$$

We will use these estimated incidence rates, specifically the estimate of the incidence rate in the cell corresponding to non-smoking AUP-users, to compare results from models that can be on different scales than the generated data. For instance, a logistic regression model applied to data that was generated additively will have the limitation that there is not a clear way to compare the odds ratio estimates from the regression model to the true parameter estimates, which will be in terms of differences in attributable risk. By using the regression models to estimate incidence rates in each cell we are able to compare the results from the regression model to the generated data.

While it is not informative to compare parameter estimates across different scales directly, it can be informative to compare the interpretations of those estimates to one another. For example, if a logistic regression model estimates that AUP use is less harmful than moderate cigarette use, but is fit to data that is truly additive and in which AUP use is actually more harmful, we can determine that the violation of scale assumptions leads to misleading results beyond that which arise when comparing odds ratios to risk differences. We will investigate scenarios in which inference can be misleading when logistic regression is applied to additively generated data, comparing whether the estimated odds ratios for AUP use are greater or less than those estimated for moderate and heavy cigarette use with whether the true risk differences lead to similar conclusions.

Tables 8 through 10 summarize all of the data sets that we will use, with chosen true incidence rates as well as the true conditional AUP effect sizes for each model under the various model scale assumptions: attributable risk (AR), relative risk (RR), and odds ratio (OR).

Table 8: Example 3, Summary of Additive Data Sets

	Cig	AUP	IR	AR	RR	OR
Data set 1	None	No	0.150			
		Yes	0.100	-0.05	0.67	0.63
	Mod	N	0.225			
		Y	0.175	-0.05	0.78	0.73
	Heavy	N	0.300			
		Y	0.250	-0.05	0.83	0.78
Data set 2	None	No	0.150			
		Yes	0.150	0	1	1
	Mod	N	0.225			
		Y	0.225	0	1	1
	Heavy	N	0.300			
		Y	0.300	0	1	1
Data set 3	None	No	0.150			
		Yes	0.200	0.050	1.33	1.42
	Mod	N	0.225			
		Y	0.275	0.050	1.22	1.31
	Heavy	N	0.300			
		Y	0.350	0.050	1.17	1.26
Data set 4	None	No	0.150			
		Yes	0.275	0.125	1.83	2.15
	Mod	N	0.225			
		Y	0.350	0.125	1.56	1.85
	Heavy	N	0.300			
		Y	0.425	0.125	1.42	1.72
Data set 5	None	No	0.150			
		Yes	0.350	0.200	2.33	3.05
	Mod	N	0.225			
		Y	0.425	0.200	1.89	2.55
	Heavy	N	0.300			
		Y	0.500	0.200	1.67	2.33

Table 9: Example 3, Summary of Multiplicative (Relative Risk) Data Sets

	Cig	AUP	IR	AR	RR	OR
Data set 6	None	No	0.150			
		Yes	0.100	-0.050	0.67	0.63
	Mod	N	0.225			
		Y	0.151	-0.074	0.67	0.61
	Heavy	N	0.300			
		Y	0.201	-0.099	0.67	0.59
Data set 7	None	No	0.150			
		Yes	0.150	0	1	1
	Mod	N	0.225			
		Y	0.225	0	1	1
	Heavy	N	0.300			
		Y	0.300	0	1	1
Data set 8	None	No	0.150			
		Yes	0.200	0.050	1.33	1.42
	Mod	N	0.225			
		Y	0.299	0.074	1.33	1.47
	Heavy	N	0.300			
		Y	0.399	0.099	1.33	1.55
Data set 9	None	No	0.150			
		Yes	0.275	0.125	1.83	2.15
	Mod	N	0.225			
		Y	0.412	0.187	1.83	2.41
	Heavy	N	0.300			
		Y	0.549	0.249	1.83	2.84
Data set 10	None	No	0.150			
		Yes	0.350	0.200	2.33	3.05
	Mod	N	0.225			
		Y	0.524	0.299	2.33	3.79
	Heavy	N	0.300			
		Y	0.699	0.399	2.33	5.42

Table 10: Example 3, Summary of Multiplicative (Odds Ratio) Data Sets

	Cig	AUP	IR	AR	RR	OR
Data set 11	None	No	0.150			
		Yes	0.100	-0.050	0.67	0.63
	Mod	N	0.225			
		Y	0.155	-0.070	0.69	0.63
	Heavy	N	0.300			
		Y	0.213	-0.087	0.71	0.63
Data set 12	None	No	0.150			
		Yes	0.150	0	1	1
	Mod	N	0.225			
		Y	0.225	0	1	1
	Heavy	N	0.300			
		Y	0.300	0	1	1
Data set 13	None	No	0.150			
		Yes	0.200	0.050	1.33	1.42
	Mod	N	0.225			
		Y	0.292	0.067	1.30	1.42
	Heavy	N	0.300			
		Y	0.378	0.078	1.26	1.42
Data set 14	None	No	0.150			
		Yes	0.275	0.125	1.83	2.15
	Mod	N	0.225			
		Y	0.384	0.159	1.71	2.15
	Heavy	N	0.300			
		Y	0.480	0.180	1.60	2.15
Data set 15	None	No	0.150			
		Yes	0.350	0.200	2.33	3.05
	Mod	N	0.225			
		Y	0.470	0.245	2.09	3.05
	Heavy	N	0.300			
		Y	0.567	0.267	1.89	3.06 [a]

[a] difference due to rounding

Since logistic regression is commonly used to analyze dichotomous outcome data our primary focus will be on examining logistic regression models on data that is not truly multiplicative on the odds ratio scale, in the presence and absence of sparse data. It is for this reason that we elected to use a relatively prevalent outcome variable, as the additive model serves as a reasonable approximation of the logistic model for probabilities that are not near 0 or 1 [19].

We will generate some additional data sets not listed in Tables 8 through 10 in order to identify whether there are scenarios in which violated scale assumptions lead to misleading inference. These will focus on additively generated data sets to which logistic regression are applied, although we will also briefly examine the consequences of applying linear regression to multiplicative data. We will seek to determine whether it is possible for the direction of inference to be reversed when modeling sparse data on an incorrect scale. One example would be if AUP use if truly more harmful than

moderate cigarette smoking on an additive scale, but logistic regression estimates AUP use to be less harmful.

We will also extend these examples by creating data sets in which age is included as a covariate and is associated with the outcome. In these data sets age will take the values 20, 25, 30, ..., 55, 60, 65 (for coding purposes these will be represented as 0, 1, 2, ..., 8, 9, 10). We will first generate age such that it is uniformly distributed independent of tobacco status. Subsequently we will also generate data sets where age is dependent on tobacco status, specifically that users of AUP products or cigarettes will tend to be more likely on the younger side of the age distribution while age of tobacco abstainers will be distributed uniformly. Age will be treated as a continuous variable with an association with outcome that is linear on the scale of the generated data set (e.g., linear according to the logistic scale when one of the OR-scale data sets is used). Age will be modeled as a linear continuous variable in each regression model (again, according to the scale of the regression model), with no interaction between age and either tobacco variable. The purpose of including a third covariate into these models is that the interaction models A, C, and F are all fully parameterized models (six coefficients for a model with six unique combinations of covariates), and as such are effectively reparameterizations of one another. By including age and fitting models as described below the interaction models are no longer reparameterizations of main effect models using a different scale. The regression models for these examples are provided below.

$$\begin{aligned} X_1, X_2, X_3 &= \text{as defined previously} \\ X_4 &= \text{age} \end{aligned}$$

Model A′: $P(Dz|X_1 = x_1, X_2 = x_2, X_3 = x_3, X_4 = x_4) = \gamma_0 + \gamma_1 x_1 + \gamma_2 x_2 + \gamma_3 x_3 + \gamma_4 x_1 x_3 + \gamma_5 x_2 x_3 + \gamma_6 x_4$

Model B′: $P = \gamma_0^* + \gamma_1^* x_1 + \gamma_2^* x_2 + \gamma_3^* x_3 + \gamma_4^* x_4$

Model C′: $\text{logit}(P) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_1 x_3 + \beta_5 x_2 x_3 + \beta_6 x_4$

Model D′: $\text{logit}(P) = \beta_0^* + \beta_1^* x_1 + \beta_2^* x_2 + \beta_3^* x_3 + \beta_4^* x_4$

Incidence rates will be calculated as described above, with the added restriction that they will be calculated for subjects in the age 20 category. When including age we will only consider data sets and models that are on the additive or multiplicative (OR) scale once age is included.

Since we hypothesize that sparse data will lead to a lack of power to detect interaction it will be of interest to determine how many observations are required to detect interaction in data sets in which we know it is present. We will check this in two ways, by increasing the number of observations in the sparse data cell corresponding to AUP use but no cigarette use as well as by increasing the total number of observations in the study proportionally over all six cigarette/AUP combinations. In both cases we will increase the number of observations until the LRT for significance of interaction terms is less than 0.05. This will be done for additive data sets (not including age) to which logistic regression is applied and to multiplicative data sets (including age and interaction between AUPs and cigarettes) to which logistic regression is applied. For the multiplicative data sets we will use three different interaction effect sizes between AUP use and cigarette use: AUP odds ratio is 22% more harmful among smokers than non-smokers, 28% higher, and 35% higher.

Example 4: Exploration of Coarse Categorization and Sparsity in Cross-Sectional Cohort Study

This example also uses the basic framework of the PATH study, generating data sets consisting of 40,000 subjects according to the same mechanism as described in Example 3. However, the regression models that will be fit to the data will use a coarser categorization for cigarette use than what was used to generate the data, treating cigarette use as a dichotomous variable in which moderate and heavy smoking are collapsed into one category. We will also focus solely on data sets that are generated multiplicatively (odds ratio scale) and additively, and on logistic and linear regression. These regression models are described below:

$$\begin{aligned} X_3 &= \text{as defined previously} \\ X_5 &= \begin{cases} 1 \text{ if } X_1 = 1 \text{ or } X_2 = 1 \\ 0 \text{ otherwise} \end{cases} \end{aligned}$$

$$\begin{aligned} \text{Model A: } P(Dz|X_5 = x_5, X_3 = x_3) &= \gamma_0 + \gamma_1 x_5 + \gamma_2 x_3 + \gamma_3 x_3 x_5 \\ \text{Model B: } P &= \gamma_0^* + \gamma_1^* x_5 + \gamma_2^* x_3 \\ \text{Model C: } \text{logit}(P) &= \beta_0 + \beta_1 x_5 + \beta_2 x_3 + \beta_3 x_5 x_3 \\ \text{Model D: } \text{logit}(P) &= \beta_0^* + \beta_1^* x_5 + \beta_2^* x_3 \end{aligned}$$

We have also altered the 3x2 frequencies of cigarette use versus AUP use for the sparse and non-sparse scenarios, with the aim of creating data sets in which AUP users who smoke tend to be moderate smokers, while smokers who do not use an AUP tend to be heavy smokers. These frequencies are presented in Table 11

Table 11: Example 4, AUP Status by Cigarette Status for Sparse and Non-sparse Data Sets

	Sparse		Non-sparse	
	AUP		AUP	
Cigarettes	Yes	No	Yes	No
Heavy	200 (0.5%)	5200 (13.0%)	640 (1.6%)	16906 (42.3%)
Moderate	1400 (3.5%)	1400 (3.5%)	4560 (11.4%)	4560 (11.4%)
None	200 (0.5%)	31600 (79.0%)	6667 (16.7%)	6666 (16.7%)

We will also briefly examine examples of sparsity in which we reverse the relationship between heavy/moderate cigarette use and AUP use, as shown in Table 12.

Table 12: Example 4, AUP Status by Cigarette Status for Sparse Data Sets with Reversed Association

	AUP	
Cigarettes	Yes	No
Heavy	1400 (3.5%)	1400 (3.5%)
Moderate	200 (0.5%)	5200 (13.0%)
None	200 (0.5%)	31600 (79.0%)

We are interested in investigating what issues arise when a single value of a covariate contains multiple "true" values (e.g., moderate and heavy smokers have different risks but have been grouped together) and the distribution of those "true" values is different for different values of a second covariate (e.g., AUP users who do smoke tend to be lower intensity smokers than non-AUP users). This is why the non-sparse data is not balanced as in Example 3; in order to compare the effect of sparsity on statistical analysis in this situation, we need to have unbalanced AUP status by cigarette status. We are also not as concerned with the issue of incorrect scale assumptions which was the focus of Example 3. We believe that issues stemming from coarse categorizations can arise regardless of scale, and will focus on what consequences arise even when correct scale assumptions are made.

We will follow a similar procedure as in Example 3, first fitting models that include interaction terms, testing the interaction terms for significance using the LRT at the $\alpha = 0.05$ level, and if not significant reporting results from the reduced model that contains only main effects. Since we are interested in issues related to coarse categorizations when scale assumptions have not been violated, we will evaluate the accuracy of the regression models by comparing the parameter estimates to the true parameter values in each data set, rather than comparing estimated and true incidence rates. We will focus on the parameter estimates for AUP use among non-smokers, a category that is not affected by the decision to combine moderate and heavy smokers and hence can be appropriately compared to the true parameter value for that category of tobacco use.

Finally, we will briefly extend this example by considering data sets in which cigarette use is split into four categories (no use, light use, moderate use, and heavy use) but the regression models dichotomize the variable (no/light use vs. moderate/heavy use). This is done so that we can examine cases where the parameter which we are most interested in (AUP use among those at the lowest intensity smoking category) is affected directly by coarse categorization. Table 13 presents the frequencies of cigarette status by AUP use for this extension.

Table 13: Example 4, AUP Status by Four Level Cigarette Status for Sparse Data Sets

	AUP	
Cigarettes	Yes	No
Heavy	200 (0.5%)	5000 (12.5%)
Moderate	1000 (2.5%)	4300 (10.8%)
Light	1000 (2.5%)	4300 (10.8%)
None	200 (0.5%)	24000 (60.0%)

Example 5: Impact of Sparsity in Longitudinal Data

In our last example we aim to examine a more complicated scenario stemming from the PATH framework. The previous examples had ignored the possibility of tobacco use changing over time, and our underlying data generation only took intensity of tobacco use into consideration, not duration. While this is appropriate as a means of replicating cross-sectional or retrospective studies, a longitudinal cohort study such as the PATH study will be able to monitor tobacco usage patterns over time. In order to simulate this, we will define tobacco intensity similarly to previous examples; no smoking vs. moderate smoking vs. heavy smoking and AUP user vs. AUP non-user. However, we will now allow subjects' tobacco use statuses to change by generating this status at a time 0 (year 0 of study) and a time 1 (year 10 of study). We will consider each subject to continue their tobacco use for the 10 years following the recording of that status. Outcomes will then be assessed at time 2 (year 20).

There are 36 different possible combinations of tobacco use statuses at the two time points. For simplicity we will not generate any subjects with 20 of those statuses, leaving 16 to consider. These were combinations that we believed would be rarely observed. Combinations not generated include progressing from heavy cigarette smoking to no tobacco use, quitting AUP use, and switching from being an AUP user and moderate cigarette smoker to only being a heavy cigarette smoker. The numbers of subjects for each combination is presented in Table 14.

Table 14: Example 5, Generated Tobacco Use Over Time

	Status at Year 10					
Status at Year 0	None	Mod. cig.	Hvy. cig.	AUP	Mod. cig. + AUP	Hvy. cig + AUP
None	27900	1200	0	30	0	0
Mod. cig.	930	4160	0	200	15	0
Hvy. cig.	0	1140	4170	0	40	10
AUP	0	0	0	70	0	0
Mod. + AUP	0	0	0	70	25	0
Hvy. + AUP	0	0	0	0	20	20

In this data set 12,000 (30%) of subjects used cigarettes at some point during the study and 500 (1.25%) used the AUP of interest. Of the 12,000 who smoked, 4,200 (35%) smoked moderately throughout, 4,200 (35%) smoked heavily throughout, and 1,200 (10%) were in each of the following categories: began smoking (at a moderate level) after 10 years on study, quit smoking (from a moderate level) after 10 years, and decreased from heavy to moderate smoking after 10 years. Of the 4,200 who were continuous moderate smokers, 40 (1.0%) used AUP at some point. Of the 4,200 who were continuous heavy smokers, 30 (0.7%) used the AUP. 100 (0.4%) of the 28,000 non-smokers, 60 (5.0%) of the 1,200 decreased intensity smokers, and 270 (22.5%) of those who quit smoking used the AUP at some point during the study period. These frequencies mimic a study in which AUP use is observed relatively rarely, but is especially rarely observed among people who do not have a history of cigarette use. Of the 500 who used the AUP at some point during the study, 200 (40%) of those were people who began using the AUP after quitting moderate smoking.

As a comparison we also used a data set without sparsity. All cell combinations that were not observed in our original data set were also not observed in the non-sparse data set, and all cell combinations that were observed were set to 3000 observations. This resulted in a data set

consisting of 48,000 total subjects.

We used myocardial infarction (MI) as our hypothetical outcome for this analysis. We fixed the observed rates of MI for the five discrete cigarette use combinations (continuous moderate, continuous heavy, former moderate, decreased, and new moderate), two AUP use combinations (continuous AUP, new AUP), and for those who entirely abstained from tobacco. These observed rates of MI were set to be 0.0083 for tobacco abstainers, 0.011 for former moderate smokers, 0.0275 for new moderate smokers, 0.035 for continuous moderate smokers, 0.0425 for decreased intensity smokers, and 0.055 for continuous heavy smokers. Observed rates for AUP users were varied to represent different possible AUP effect sizes: no effect (0.0083 for both continuous and new AUP users), not as harmful as moderate cigarette use (0.015 for new AUP users, 0.019 for continuous AUP users), same effect as moderate cigarette use (0.0275 for new, 0.035 for continuous), effect between that of moderate and heavy cigarette use (0.04 for new, 0.051 for continuous), and more harmful effect than for heavy cigarette use (0.05 for new, 0.065 for continuous). Even if these relationships are unlikely for MI, they can serve to represent other health outcomes where it is more reasonable to believe that AUP use could be more harmful than cigarette use, such as forms of oral cancer when the AUP is a form of smokeless tobacco. The observed incidence rates for combinations of cigarette and AUP use were calculated assuming a multiplicative association between cigarette and AUP use. As an example, if the observed AUP incidence rates are 0.015 for new users and 0.019 for continuous users, the observed incidence rate for continuous moderate smokers who also were continuous AUP users will be equal to the inverse logit of the sum of the β coefficients corresponding to the baseline risk (β_0), risk associated with continuous moderate smoking (β_{mod20}), and risk associated with continuous AUP use (β_{aup20}):

$$\begin{aligned} IR &= \text{expit}(\beta_0 + \beta_{mod20} + \beta_{aup20}) \\ &= \text{expit}\Big(\text{logit}(0.0083) + \big(\text{logit}(0.035) - \text{logit}(0.0083)\big) + \big(\text{logit}(0.019) - \text{logit}(0.0083)\big)\Big) \\ &= 0.077 \end{aligned}$$

Figures 1, 2, and 3 show the observed incidence rates for selected AUP and moderate smoking combinations for three of the AUP effect sizes that were used.

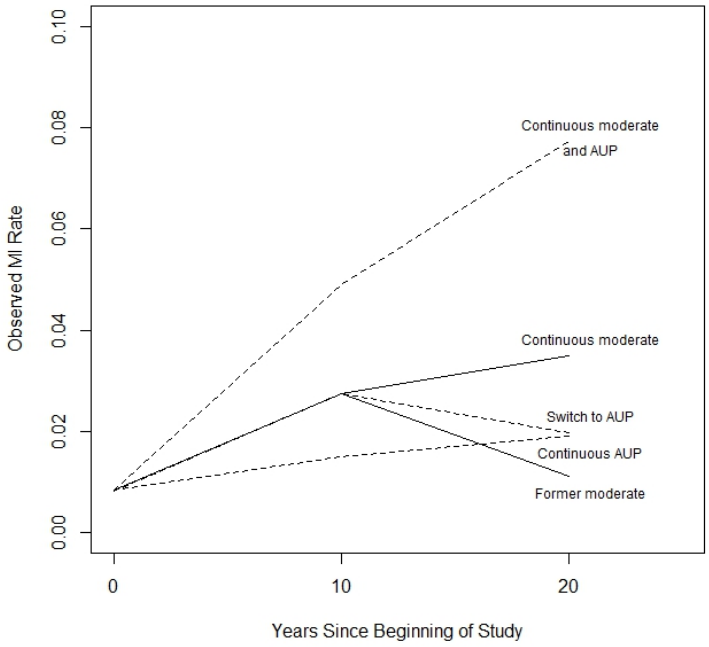

Figure 1: Selected Incidence Rates When AUP Less Harmful than Moderate Cig Use

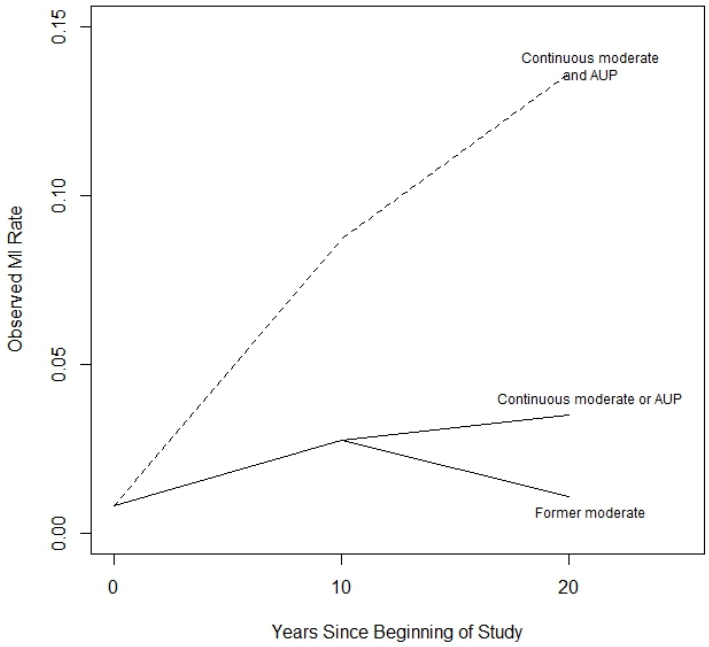

Figure 2: Selected Incidence Rates When AUP Same as Moderate Cig Use

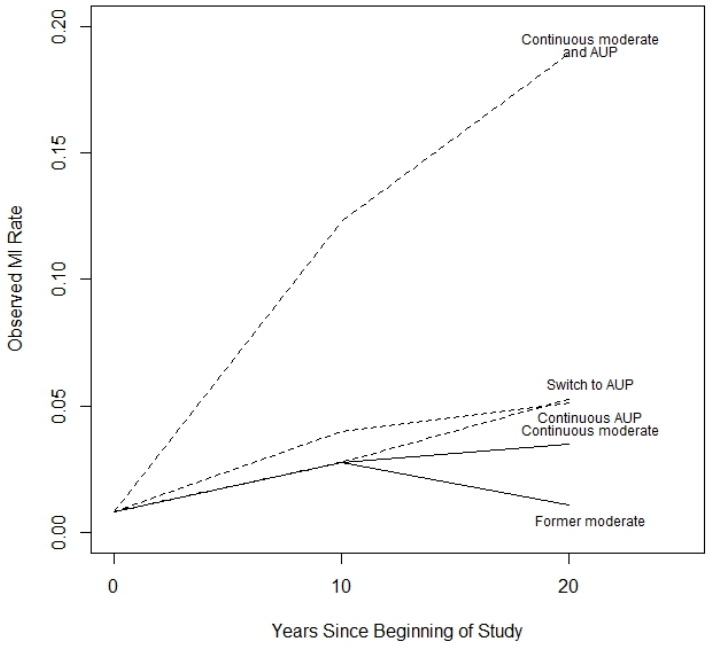

Figure 3: Selected Incidence Rates When AUP More Harmful than Moderate Cig Use

For modeling of the data from these hypothetical PATH studies we used two different covariate categorizations. In the first scheme we dichotomized both cigarette and AUP use as ever used vs. never used. In the second we trichotomized cigarette use as currently use vs. formerly used vs. never used while keeping AUP use dichotomous. These schemes are meant to replicate the categorization that is often necessary when attempting to analyze longitudinal tobacco use, although there are many other methods that could be used as well. We performed logistic regression using these covariate categorizations, and compared the odds ratio estimates from those regression models to what we knew to be the true effects of AUP and cigarette use.

As an extension to this example we also generated some data sets in which subjects experienced mortality during the 20 years of our hypothetical study. This was done by dropping 10% of the subjects that experienced MI and 1% of those that did not, proportionally across all tobacco use categories. We then implemented the same logistic regression models in order to examine what impact the loss to follow-up could have on an analysis that simply excluded such observations.

3 Results

Example 1

Example 1 uses a hypothetical data set from a case-control study with 360 controls and 180 cases. Six of the eight combinations of case/AUP/cigarette status have been fixed as shown in Table 4. This results in the odds ratio estimating the association between AUP use and esophageal cancer to be $\frac{45(120)}{45(100)} = 1.2$ and the unadjusted marginal odds ratio estimating the association between cigarette use and esophageal cancer to be $\frac{(120+100)(45+45)}{(a+b)(45+45)} = \frac{220(90)}{(a+b)(90)} = 1.57$, since the fixed number of cases and controls and specific cell counts in Table 4 ensure that $a + b = 140$. In the following tables we present six combinations of a and b that satisfy Scenarios 1-6 as described previously.

Table 15: Example 1, Scenario 1 Cell Counts
Marginal AUP OR > Conditional, Cigarette ORs

	Cases		Controls	
	AUP		AUP	
Cigarette status	Yes	No	Yes	No
Yes	45	45	20	120
No	45	45	100	120

In Scenario 1 (Table 15), we obtain a marginal estimate of the AUP odds ratio of $\frac{(120+120)(45+45)}{(100+20)(45+45)} = 2.0$. This estimate implies an effect that is quite a bit more harmful than the estimate of AUP effect that is limited to non-smokers. It is also a larger estimate than that which we estimated for cigarette use.

Table 16: Example 1, Scenario 2 Cell Counts
Marginal AUP OR Between Conditional, Cigarette ORs

	Cases		Controls	
	AUP		AUP	
Cigarette status	Yes	No	Yes	No
Yes	45	45	50	90
No	45	45	100	120

By redistributing the non-smoking controls to have 50 AUP users rather than 20 (shown in Table 16) the marginal odds ratio is now estimated to be $\frac{(120+90)(45+45)}{(100+50)(45+45)} = 1.4$. This satisfies our conditions for Scenario 2, as this odds ratio is between the marginal cigarette odds ratio of 1.57 and the conditional odds ratio for AUP use among non-smokers of 1.2.

Table 17: Example 1, Scenario 3 Cell Counts
Marginal AUP OR Equal to To Conditional AUP OR

	Cases		Controls	
	AUP		AUP	
Cigarette status	Yes	No	Yes	No
Yes	45	45	64	76
No	45	45	100	120

The marginal AUP odds ratio from Table 17 is estimated to be $\frac{(120+76)(45+45)}{(100+64)(45+45)} = 1.2$. This satisfies Scenario 3, the trivial case in which the marginal and main effect-only AUP odds ratios are the same. In this case there would be no evidence of interaction between cigarette use and AUP use.

Table 18: Example 1, Scenario 4 Cell Counts
Marginal AUP OR Between No Effect and Conditional OR

	Cases		Controls	
	AUP		AUP	
Cigarette status	Yes	No	Yes	No
Yes	45	45	70	70
No	45	45	100	120

The cell frequencies in Table 18 meet the conditions for Scenario 4 by resulting in a marginal odds ratio estimate for AUP use that is $\frac{(120+70)(45+45)}{(100+70)(45+45)} = 1.12$, or a slightly harmful association that is of lesser magnitude than the conditional odds ratio for AUP use given no cigarette use.

Table 19: Example 1, Scenario 5 Cell Counts
Marginal AUP OR = 1

	Cases		Controls	
	AUP		AUP	
Cigarette status	Yes	No	Yes	No
Yes	45	45	80	60
No	45	45	100	120

The data set reported in Table 19 results in a marginal odds ratio estimate for AUP use of $\frac{(120+60)(45+45)}{(100+80)(45+45)} = 1$. This meets Scenario 5, as this effect is neutral, despite the odds ratio for AUP use among non-smokers showing a higher risk of esophageal cancer for AUP users.

Table 20: Example 1, Scenario 6 Cell Counts
Marginal AUP OR < 1

	Cases		Controls	
	AUP		AUP	
Cigarette status	Yes	No	Yes	No
Yes	45	45	100	40
No	45	45	100	120

Finally, Table 20 shows that it is possible to satisfy Scenario 6 and have the marginal AUP odds ratio be in the opposite direction of both the marginal cigarette odds ratios and the conditional odds ratio for AUP use among non-smokers, as $\frac{(120+40)(45+45)}{(100+100)(45+45)} = 0.8$. These six scenarios demonstrate that it is possible to observe data sets in which the marginal odds ratio for AUP use can range from being in the opposite direction of both the conditional odds ratio for AUP use among non-smokers and the marginal odds ratio for cigarette use to being more extreme in the same direction as both those estimates.

Example 2
Marginal univariate odds ratios for cigarette and moist snuff use can be calculated using the frequencies previously reported in Tables 5 and 6. These estimates are $\frac{90(154)}{41(166)} = 2.04$ for cigarette use and $\frac{20(255)}{65(112)} = 0.70$ for snuff use. These calculations assume no interaction between the two tobacco products and, as seen in Example 1, violations of this assumption can have large implications for parameter estimation. Table 21 presents a 2x2x2 table which shows what is known about the joint relationship between OOSCC, cigarette, and moist snuff statuses based on the published results.

Table 21: Example 2, Moist Snuff Use, Cigarette Use, and OOSCC

	Cases			Controls			Total
	Snuff			Snuff			
Cigarette use	Ever	Never	Total	Ever	Never	Total	
Ever	a	b	90	e	f	166	256
Never	c	d	41	g	h	154	195
Total	20	111 [a]	131	65	255	320	451

[a] 1 subject missing cigarette data removed from this cell

Cells a through h are unknown, although we do know pairwise sums, such as $a+b = 90$. Because these sums are known, the value of one of a, b, c, or d fully determines the values of the other three, and the same is true for the grouping of e, f, g, and h. For example, if we know that $d = 25$, then $b = 111 - 25 = 86$, $c = 41 - 25 = 16$, and $a = 131 - 25 - 86 - 16 = 4$. So, by fixing one of a, b, c, or d and one of e, f, g, or h we can determine all the cell values in Table 21. We have chosen to control the two cells corresponding to non-smoker/non-snuff user (d and h), but this is an arbitrary decision; any independent pair would suffice. In order to maintain non-negative values for all cells in the table, d must be between 21 and 41, and h between 89 and 154. This results in 21 possible arrangements of the cases and 66 arrangements of the controls, for 1386 possible combinations that still satisfy the known marginal odds ratios. However, in order to calculate odds ratios for moist snuff use within smoking levels we must have non-zero values in all cells, so we can further restrict

the range of d and h to be 22 to 40 and 90 to 153, respectively. This leaves us with 1216 possibilities to consider.

Figures 4 and 5 show the odds ratio and log odds ratio estimates for snuff use among non-smokers for all possible values of h and 5 different values for d. The smallest possible main effect odds ratio is 0.035, which would be observed if $d = 40$ and $h = 90$. This is the largest possible value for d, and smallest possible for h (given the no zero cell restriction). Conversely, a data set with the smallest possible value for d ($d = 22$) and largest possible for h ($h = 153$) results in the maximum main effect odds ratio of 132.1. Figures 6 and 7 show the full range of possible values and resulting odds ratios and log odds ratios, rather than only for selected values of d as in Figures 4 and 5.

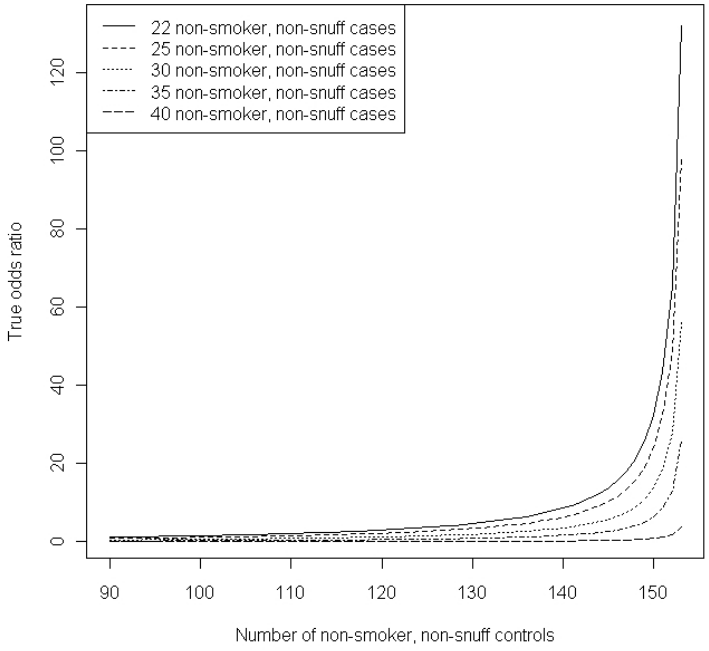

Figure 4: Possible Main Effect Odds Ratios

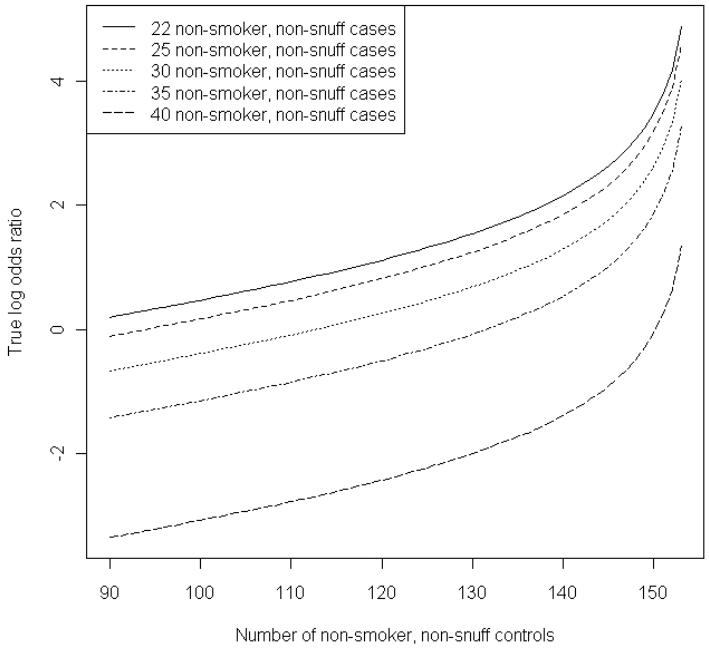

Figure 5: Possible Main Effect Log ORs

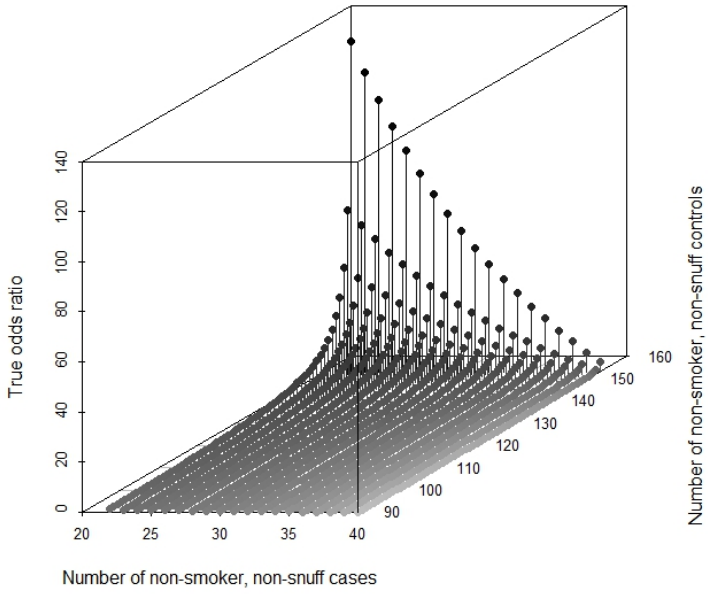

Figure 6: All Possible Main Effect Odds Ratios

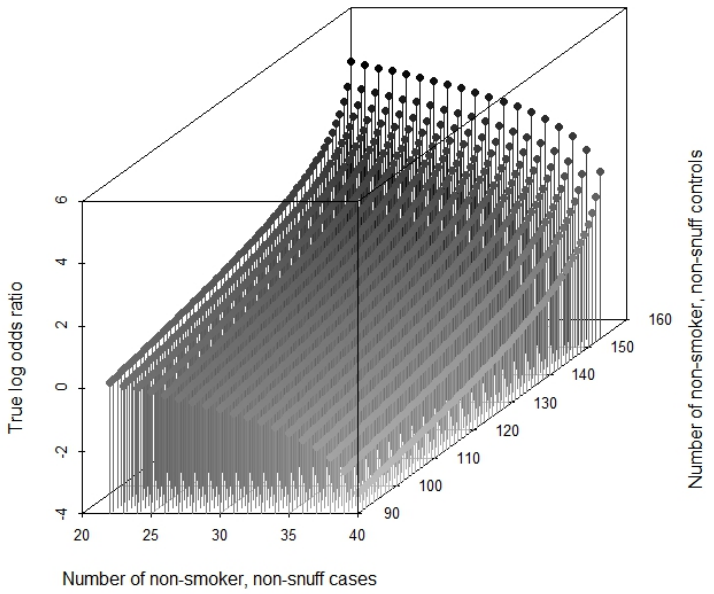

Figure 7: All Possible Main Effect Log ORs

We also tested whether or not there was statistically significant evidence of interaction in these regression models. In 786 (64.6%) of 1216 data sets the LRT p-value was less than 0.05 and would have provided statistically significant evidence of interaction. We found that 252 (20.7%) of data sets had significant evidence of interaction with an odds ratio for the association between moist snuff use and OOSCC greater than 1 among non-smokers, and 534 (43.9%) with an odds ratio less than 1.

While we have only reported odds ratio estimates for moist snuff use among non-smokers it is also informative to look at the estimates among smokers. Table 22 presents a comparison of the odds ratio estimates for moist snuff use among non-smokers versus smokers, specifically whether or not that odds ratio is estimated to be greater than or less than 1.

Table 22: Example 2, Summary of Conditional Moist Snuff Odds Ratio Estimates

		OR(Snuff|Cig=Ever)		
		< 1	> 1	Total
OR(Snuff|Cig=Never)	< 1	192 (15.8%)	683 (56.2%)	875
	> 1	329 (27.1%)	12 (1.0%)	341
Total		521	695	1216

56.2% of the data sets would have estimated the odds ratio for snuff use given non-smoker status to be harmful while the odds ratio given being a smoker would be protective, while 27.1% would be the opposite. 15.8% would have estimated the odds ratios given both smoker statuses to be protective, while 1.0% would estimate both odds ratios to be harmful. This last scenario is

a particular concern since we know that the marginal odds ratio estimate was 0.70, or protective, meaning that it is possible to observe a data set in which the marginal moist snuff use odds ratio is in the opposite direction of *both* of the conditional moist snuff use odds ratios. It is important to note that these odds ratio estimates solely reflect the effect size and not any measure of statistical significance. Of the 12 data sets that lead to harmful odds ratio estimates for moist snuff use among both smokers and non-smokers, none were sufficiently extreme to conclude that there was evidence of interaction (all LRT $p > 0.298$). These data sets tended to have very few subjects who used both snuff and cigarettes as all 12 data sets had fewer than 15 subjects in that group, and no more than 5 cases and 9 controls.

Example 3

Table 23 provides a reference guide for the results that will be presented from the PATH examples. Each cell in the table refers to one data set.

Table 23: Example 3, Reference Guide

	AUP, cigarettes only		Age included	
True	Sparse	Non-sparse	Sparse	Non-sparse
Additive				
AR = -0.050	Tbl 24, rows 1, 6, 11	Tbl 27, rows 1, 6, 11	Tbl 30, rows 1, 6	Tbl 33, rows 1, 6
AR = 0.000 [a]	Tbl 24, rows 2, 7, 12	Tbl 27, rows 2, 7, 12	Tbl 30, rows 2, 7	Tbl 33, rows 2, 7
AR = 0.050	Tbl 24, rows 3, 8, 13	Tbl 27, rows 3, 8, 13	Tbl 30, rows 3, 8	Tbl 33, rows 2, 7
AR = 0.125	Tbl 24, rows 4, 9, 14	Tbl 27, rows 4, 9, 14	Tbl 30, rows 4, 9	Tbl 33, rows 2, 7
AR = 0.200	Tbl 24, rows 5, 10, 15	Tbl 27, rows 5, 10, 15	Tbl 30, rows 5, 10	Tbl 33, rows 5, 10
Multiplicative (OR)				
OR = 0.63	Tbl 25, rows 1, 6, 11	Tbl 28, rows 1, 6, 11	Tbl 31, rows 1, 6	Tbl 34, rows 1, 6
OR = 1.00 [a]	Tbl 24, rows 2, 7, 12	Tbl 27, rows 2, 7, 12	Tbl 30, rows 2, 7	Tbl 33, rows 2, 7
OR = 1.42	Tbl 25, rows 3, 8, 13	Tbl 28, rows 3, 8, 13	Tbl 31, rows 3, 8	Tbl 34, rows 2, 7
OR = 2.15	Tbl 25, rows 4, 9, 14	Tbl 28, rows 4, 9, 14	Tbl 31, rows 4, 9	Tbl 34, rows 2, 7
OR = 3.05	Tbl 25, rows 5, 10, 15	Tbl 28, rows 5, 10, 15	Tbl 31, rows 5, 10	Tbl 34, rows 5, 10
Multiplicative (OR) with interaction				
OR = 0.63			Tbl 32, rows 1, 5	Tbl 35, rows 1, 5
OR = 1.42			Tbl 32, rows 2, 6	Tbl 35, rows 2, 6
OR = 2.15			Tbl 32, rows 3, 7	Tbl 35, rows 3, 7
OR = 3.05			Tbl 32, rows 4, 8	Tbl 35, rows 4, 8
Multiplicative (RR)				
RR = 0.67	Tbl 26, rows 1, 6, 11	Tbl 29, rows 1, 6, 11		
RR = 1.00 [a]	Tbl 24, rows 2, 7, 12	Tbl 27, rows 2, 7, 12		
RR = 1.33	Tbl 26, rows 3, 8, 13	Tbl 29, rows 3, 8, 13		
RR = 1.83	Tbl 26, rows 4, 9, 14	Tbl 29, rows 4, 9, 14		
RR = 2.33	Tbl 26, rows 5, 10, 15	Tbl 29, rows 5, 10, 15		

[a] AR=0, OR=1, and RR=1 are identical

Tables 24, 25, and 26 present the results from the 13 sparse data sets used to assess issues related to scale. Each row in these tables represents a combination of data set and regression model; the first row in Table 24 represents an additive data set where AUP use is associated with a 0.05 reduction in attributable risk, to which linear regression is applied. This is represented in

the column headed "True AR", for true attributable risk. "True IR" refers to the true incidence rate for the outcome among non-smoker AUP-users, or what we have been referring to as the main effect. In the first row this incidence rate is 0.100. "LRT p-value" provides the p-value for the test comparing the regression model including interaction to the one without. Here that p-value is 1, indicating no evidence of an interaction between AUP use and smoking in the data set (since we set the data set up such that there was no interaction on the additive scale and performed linear regression in this case, this result is not surprising). The "Used?" column indicated whether the interaction or reduced model was used (this was determined by whether the LRT p-value was significant at the 0.05-level), although the estimated incidence rates have been provided for both models. "Est. IR" reports the estimated incidence rate as determined by that regression model, and "Δ" reports the difference between the estimated incidence rate and the true incidence rate.

Table 24: Example 3, Main Effect Incidence Rate Estimates in Additive, Sparse Data Sets

Linear Regression								
			Reduced Model (Model B)			Interaction Model (Model A)		
True AR	True IR	LRT p-value	Used?	Est. IR	Δ	Used?	Est. IR	Δ
-0.050	0.100	1.000	x	**0.100**	**0.000**		0.100	0.000
0.000	0.150	1.000	x	**0.150**	**0.000**		0.150	0.000
0.050	0.200	1.000	x	**0.200**	**0.000**		0.200	0.000
0.125	0.275	1.000	x	**0.275**	**0.000**		0.275	0.000
0.200	0.350	1.000	x	**0.350**	**0.000**		0.350	0.000
Logistic Regression								
			Reduced Model (Model D)			Interaction Model (Model C)		
True AR	True IR	LRT p-value	Used?	Est. IR	Δ	Used?	Est. IR	Δ
-0.050	0.100	0.673	x	**0.116**	**0.016**		0.100	0.000
0.000	0.150	1.000	x	**0.150**	**0.000**		0.150	0.000
0.050	0.200	0.821	x	**0.186**	**-0.014**		0.200	0.000
0.125	0.275	0.459	x	**0.244**	**-0.031**		0.275	0.000
0.200	0.350	0.281	x	**0.307**	**-0.043**		0.350	0.000
Relative Risk Regression								
			Reduced Model (Model F)			Interaction Model (Model E)		
True AR	True IR	LRT p-value	Used?	Est. IR	Δ	Used?	Est. IR	Δ
-0.050	0.100	0.530	x	**0.120**	**0.020**		0.100	0.000
0.000	0.150	1.000	x	**0.150**	**0.000**		0.150	0.000
0.050	0.200	0.647	x	**0.180**	**-0.020**		0.200	0.000
0.125	0.275	0.106	x	**0.225**	**-0.050**		0.275	0.000
0.200	0.350	0.007		0.271	-0.079	x	**0.350**	**0.000**
Poisson Regression								
			Reduced Model (Model F)			Interaction Model (Model E)		
True AR	True IR	LRT p-value	Used?	Est. IR	Δ	Used?	Est. IR	Δ
-0.050	0.100	0.587	x	**0.120**	**0.020**		0.100	0.000
0.000	0.150	1.000	x	**0.150**	**0.000**		0.150	0.000
0.050	0.200	0.720	x	**0.181**	**-0.019**		0.200	0.000
0.125	0.275	0.210	x	**0.227**	**-0.048**		0.275	0.000
0.200	0.350	0.039		0.273	-0.077	x	**0.350**	**0.000**

As seen in Table 24, all five linear regression models found no evidence of interaction in any of the five additive data sets ($p = 1$ for all). Again, since these data sets were set up such that there was no interaction on the attributable risk scale this is exactly what we would expect to observe. The estimated incidence rates from the linear regression models were also unbiased estimators of the true incidence rate, regardless of whether one looked at the reduced or interaction models (since the coefficients for those interaction terms in the interaction model were exactly equal to or very close to 0).

The LRT also did not find evidence of interaction on the odds ratio scale for any of the additive data sets when logistic regression was applied ($p > 0.281$ for all models). We can also see that implementing the reduced logistic model results in biased incidence rate estimates. Aside from the data set where the true attributable risk was 0 (when there was no association between AUP use and outcome) the reduced logistic regression model underestimated the magnitude of the attributable risk difference (e.g., overestimated the AUP use/no cigarette use incidence rate when AUP use was protective, underestimated when harmful). We also note that the interaction models, despite the lack of statistically significant interaction terms, generated unbiased estimations of the main effect incidence rate.

Table 25: Example 3, Main Effect Incidence Rate Estimates in Multiplicative (OR scale), Sparse Data Sets

Linear Regression								
			Reduced Model (Model B)			Interaction Model (Model A)		
True OR	True IR	LRT p-value	Used?	Est. IR	Δ	Used?	Est. IR	Δ
0.63	0.100	0.419	x	**0.075**	**-0.025**		0.100	0.000
1.42	0.200	0.658	x	**0.220**	**0.020**		0.200	0.000
2.15	0.275	0.185	x	**0.314**	**0.039**		0.275	0.000
3.05	0.350	0.085	x	**0.399**	**0.049**		0.350	0.000
Logistic Regression								
			Reduced Model (Model D)			Interaction Model (Model C)		
True OR	True IR	LRT p-value	Used?	Est. IR	Δ	Used?	Est. IR	Δ
0.63	0.100	0.999	x	**0.100**	**0.000**		0.100	0.000
1.42	0.200	0.998	x	**0.200**	**0.000**		0.200	0.000
2.15	0.275	1.000	x	**0.275**	**0.000**		0.275	0.000
3.05	0.350	1.000	x	**0.350**	**0.000**		0.350	0.000
Relative Risk Regression								
			Reduced Model (Model F)			Interaction Model (Model E)		
True OR	True IR	LRT p-value	Used?	Est. IR	Δ	Used?	Est. IR	Δ
-0.63	0.100	0.948	x	**0.105**	**0.005**		0.100	0.000
1.42	0.200	0.882	x	**0.192**	**-0.008**		0.200	0.000
2.15	0.275	0.447	x	**0.249**	**-0.026**		0.275	0.000
3.05	0.350	0.084	x	**0.300**	**-0.050**		0.350	0.000
Poisson Regression								
			Reduced Model (Model F)			Interaction Model (Model E)		
True OR	True IR	LRT p-value	Used?	Est. IR	Δ	Used?	Est. IR	Δ
-0.63	0.100	0.956	x	**0.105**	**0.005**		0.100	0.000
1.42	0.200	0.915	x	**0.192**	**-0.008**		0.200	0.000
2.15	0.275	0.588	x	**0.250**	**-0.025**		0.275	0.000
3.05	0.350	0.223	x	**0.302**	**-0.048**		0.350	0.000

Results for relative risk and Poisson regression models were fairly similar to those seen for logistic regression. Estimates of incidence rates for AUP users who did not use cigarettes were more extremely biased (further from the true incidence rate) than in the reduced logistic regression models. Related to these estimates being more extreme, LRT p-values were smaller than in the logistic regression models, and the LRT for the data set with a true attributable risk of 0.2 did pass our criteria for statistical significance ($p = 0.007$). As seen when applying logistic regression, estimates using the interaction models were unbiased estimators of the main effect incidence rate.

Table 25 presents the results of regression models on data sets that were generated on the odds ratio scale. As seen for linear regression on additive data sets, the logistic regression models show no evidence of interaction (all $p \geq 0.998$) and estimate the incidence rate for the main effect without bias.

None of the linear, relative risk, or Poisson regression models showed significant evidence of interaction (all $p > 0.08$), and, as observed previously, models which did not include interaction terms were biased in their estimation of the main effect incidence rates. We also note that, since a true odds ratio of 1 is equivalent to an attributable risk of 0 or a relative risk of 1, those rows were omitted from Tables 25 and 26.

Table 26: Example 3, Main Effect Incidence Rate Estimates in Multiplicative (RR scale), Sparse Data Sets

Linear Regression								
			Reduced Model (Model B)			Interaction Model (Model A)		
True RR	True IR	LRT p-value	Used?	Est. IR	Δ	Used?	Est. IR	Δ
0.67	0.100	0.213	x	**0.069**	**-0.031**		0.100	0.000
1.33	0.200	0.224	x	**0.231**	**0.031**		0.200	0.000
1.83	0.275	< 0.001		0.355	0.080	x	**0.275**	**0.000**
2.33	0.350	< 0.001		0.478	0.128	x	**0.350**	**0.000**
Logistic Regression								
			Reduced Model (Model D)			Interaction Model (Model C)		
True RR	True IR	LRT p-value	Used?	Est. IR	Δ	Used?	Est. IR	Δ
0.67	0.100	0.936	x	**0.096**	**-0.004**		0.100	0.000
1.33	0.200	0.857	x	**0.209**	**0.009**		0.200	0.000
1.83	0.275	0.190	x	**0.311**	**0.036**		0.275	0.000
2.33	0.350	< 0.001		0.428	0.078	x	**0.350**	**0.000**
Relative Risk Regression								
			Reduced Model (Model F)			Interaction Model (Model E)		
True RR	True IR	LRT p-value	Used?	Est. IR	Δ	Used?	Est. IR	Δ
0.67	0.100	0.999	x	**0.101**	**0.001**		0.100	0.000
1.33	0.200	1.000	x	**0.200**	**0.000**		0.200	0.000
1.83	0.275	0.999	x	**0.275**	**0.000**		0.275	0.000
2.33	0.350	1.000	x	**0.350**	**0.000**		0.350	0.000
Poisson Regression								
			Reduced Model (Model F)			Interaction Model (Model E)		
True RR	True IR	LRT p-value	Used?	Est. IR	Δ	Used?	Est. IR	Δ
0.67	0.100	0.999	x	**0.101**	**0.001**		0.100	0.000
1.33	0.200	1.000	x	**0.199**	**-0.001**		0.200	0.000
1.83	0.275	0.999	x	**0.275**	**0.000**		0.275	0.000
2.33	0.350	1.000	x	**0.349**	**-0.001**		0.350	0.000

The linear regression models for data sets with true relative risks of 0.67 and 1.33 did not exhibit significant evidence of an interaction effect ($p > 0.2$ for both). The models for more extreme relative risks (1.83 and 2.33) did show significant evidence of interaction ($p < 0.001$ for both). The incidence rate estimates based on interaction models were unbiased in their estimation of the main effect incidence rate, while the estimates from the reduced models were biased, estimating incidence rates that were further away from the non-smoking/non-AUP-user cell than in truth.

Of the logistic regression models, only the model in the highest relative risk data set (RR = 2.33) showed significant evidence of interaction ($p < 0.001$). Incidence rates in the reduced models were biased in their estimation of the main effect incidence rate, although this bias was not as great

as observed when using linear regression. The interaction models were unbiased in their estimation of the main effect incidence rate.

Relative risk regression models showed no evidence of interaction (all $p \geq 0.999$), and were unbiased regardless of whether the reduced or interaction models were implemented.

Table 27: Example 3, Main Effect Incidence Rate Estimates in Additive, Non-Sparse Data Sets

Linear Regression								
			Reduced Model (Model B)			Interaction Model (Model A)		
True AR	True IR	LRT p-value	Used?	Est. IR	Δ	Used?	Est. IR	Δ
-0.050	0.100	1.000	x	**0.100**	**0.000**		0.100	0.000
0.000	0.150	1.000	x	**0.150**	**0.000**		0.150	0.000
0.050	0.200	1.000	x	**0.200**	**0.000**		0.200	0.000
0.125	0.275	1.000	x	**0.275**	**0.000**		0.275	0.000
0.200	0.350	1.000	x	**0.350**	**0.000**		0.350	0.000
Logistic Regression								
			Reduced Model (Model D)			Interaction Model (Model C)		
True AR	True IR	LRT p-value	Used?	Est. IR	Δ	Used?	Est. IR	Δ
-0.050	0.100	0.0058		0.108	0.008	x	**0.100**	**0.000**
0.000	0.150	1.000	x	**0.150**	**0.000**		0.150	0.000
0.050	0.200	0.125	x	**0.195**	**-0.005**		0.200	0.000
0.125	0.275	< 0.001		0.264	-0.011	x	**0.275**	**0.000**
0.200	0.350	< 0.001		0.336	-0.014	x	**0.350**	**0.000**
Relative Risk Regression								
			Reduced Model (Model F)			Interaction Model (Model E)		
True AR	True IR	LRT p-value	Used?	Est. IR	Δ	Used?	Est. IR	Δ
-0.050	0.100	< 0.001		0.110	0.010	x	**0.100**	**0.000**
0.000	0.150	1.000	x	**0.150**	**0.000**		0.150	0.000
0.050	0.200	0.014		0.192	-0.008	x	**0.200**	**0.000**
0.125	0.275	< 0.001		0.259	-0.016	x	**0.275**	**0.000**
0.200	0.350	< 0.001		0.329	-0.021	x	**0.350**	**0.000**
Poisson Regression								
			Reduced Model (Model F)			Interaction Model (Model E)		
True AR	True IR	LRT p-value	Used?	Est. IR	Δ	Used?	Est. IR	Δ
-0.050	0.100	0.001		0.109	0.009	x	**0.100**	**0.000**
0.000	0.150	1.000	x	**0.150**	**0.000**		0.150	0.000
0.050	0.200	0.036		0.192	-0.008	x	**0.200**	**0.000**
0.125	0.275	< 0.001		0.259	-0.016	x	**0.275**	**0.000**
0.200	0.350	< 0.001		0.327	-0.023	x	**0.350**	**0.000**

In non-sparse additive data sets logistic and relative risk regression models were more likely to find statistically significant evidence of interaction, with only logistic regression on a data set

with true attributable risk of 0.05 not finding evidence of interaction. As in sparse data sets the interaction models were unbiased in their estimation of the main effect incidence rate, regardless of whether the correct scale regression model was implemented. In contrast to the sparse data sets, the reduced models, while still biased, are closer to the true incidence rate in the non-sparse data sets.

Table 28: Example 3, Main Effect Incidence Rate Estimates in Multiplicative (OR scale), Non-Sparse Data Sets

Linear Regression								
			Reduced Model (Model B)			Interaction Model (Model A)		
True OR	True IR	LRT p-value	Used?	Est. IR	Δ	Used?	Est. IR	Δ
0.63	0.100	< 0.001		0.091	-0.009	x	**0.100**	**0.000**
1.42	0.200	0.028		0.207	0.007	x	**0.200**	**0.000**
2.15	0.275	< 0.001		0.290	0.015	x	**0.275**	**0.000**
3.05	0.350	< 0.001		0.368	0.018	x	**0.350**	**0.000**
Logistic Regression								
			Reduced Model (Model D)			Interaction Model (Model C)		
True OR	True IR	LRT p-value	Used?	Est. IR	Δ	Used?	Est. IR	Δ
0.63	0.100	0.999	x	**0.100**	**0.000**		0.100	0.000
1.42	0.200	0.998	x	**0.200**	**0.000**		0.200	0.000
2.15	0.275	0.998	x	**0.275**	**0.000**		0.275	0.000
3.05	0.350	1.000	x	**0.350**	**0.000**		0.350	0.000
Relative Risk Regression								
			Reduced Model (Model F)			Interaction Model (Model E)		
True OR	True IR	LRT p-value	Used?	Est. IR	Δ	Used?	Est. IR	Δ
-0.63	0.100	0.513	x	**0.103**	**0.003**		0.100	0.000
1.42	0.200	0.430	x	**0.197**	**-0.003**		0.200	0.000
2.15	0.275	0.0039		0.267	-0.008	x	**0.275**	**0.000**
3.05	0.350	< 0.001		0.338	-0.012	x	**0.350**	**0.000**
Poisson Regression								
			Reduced Model (Model F)			Interaction Model (Model E)		
True OR	True IR	LRT p-value	Used?	Est. IR	Δ	Used?	Est. IR	Δ
-0.63	0.100	0.578	x	**0.102**	**0.002**		0.100	0.000
1.42	0.200	0.532	x	**0.197**	**-0.003**		0.200	0.000
2.15	0.275	0.018		0.267	-0.008	x	**0.275**	**0.000**
3.05	0.350	< 0.001		0.336	-0.014	x	**0.350**	**0.000**

Table 29: Example 3, Main Effect Incidence Rate Estimates in Multiplicative (RR scale), Non-Sparse Data Sets

Linear Regression								
			Reduced Model (Model B)			Interaction Model (Model A)		
True RR	True IR	LRT p-value	Used?	Est. IR	Δ	Used?	Est. IR	Δ
0.67	0.100	< 0.001		0.088	-0.012	x	**0.100**	**0.000**
1.33	0.200	< 0.001		0.212	0.012	x	0.200	0.000
1.83	0.275	< 0.001		0.306	0.031	x	**0.275**	**0.000**
2.33	0.350	< 0.001		0.399	0.049	x	**0.350**	**0.000**
Logistic Regression								
			Reduced Model (Model D)			Interaction Model (Model C)		
True RR	True IR	LRT p-value	Used?	Est. IR	Δ	Used?	Est. IR	Δ
0.67	0.100	0.547	x	**0.098**	**-0.002**		0.100	0.000
1.33	0.200	0.291	x	**0.203**	**0.003**		0.200	0.000
1.83	0.275	< 0.001		0.286	0.011	x	**0.275**	**0.000**
2.33	0.350	< 0.001		0.373	0.0238	x	**0.350**	**0.000**
Relative Risk Regression								
			Reduced Model (Model F)			Interaction Model (Model E)		
True RR	True IR	LRT p-value	Used?	Est. IR	Δ	Used?	Est. IR	Δ
0.67	0.100	0.994	x	**0.100**	**0.000**		0.100	0.000
1.33	0.200	0.998	x	**0.200**	**0.000**		0.200	0.000
1.83	0.275	0.999	x	**0.275**	**0.000**		0.275	0.000
2.33	0.350	0.999	x	**0.350**	**0.000**		0.350	0.000
Poisson Regression								
			Reduced Model (Model F)			Interaction Model (Model E)		
True RR	True IR	LRT p-value	Used?	Est. IR	Δ	Used?	Est. IR	Δ
0.67	0.100	0.995	x	**0.100**	**0.000**		0.100	0.000
1.33	0.200	0.998	x	**0.200**	**0.000**		0.200	0.000
1.83	0.275	0.999	x	**0.275**	**0.000**		0.275	0.000
2.33	0.350	0.999	x	**0.350**	**0.000**		0.350	0.000

The 2x2x2 frequencies of chronic bronchitis, AUP, and cigarette statuses that result in the additive data sets with AUP attributable risk of 0.05, multiplicative data sets with AUP odds ratio of 1.42, and multiplicative data sets with AUP relative risk of 1.33 are included in Tables 74 through 79 in the Appendix.

We extended this example by creating data sets that also included age as a covariate that was associated with chronic bronchitis independent of AUP and cigarette use. This was motivated by the fact that when cigarette use and AUP use were the only covariates in the model it was impossible to distinguish between data that was generated additively and data that was generated multiplicatively with interaction. If we consider truly additive data with attributable risks of 0.05,

0.075, and 0.15 for AUP use, moderate cigarette use, and heavy cigarette use, respectively, we can see that this is equivalent to a multiplicatively generated data that allows for interaction:

$$\begin{aligned}
P(\text{CB}|X_1 = x_1, X_2 = x_2, X_3 = x_3) &= \gamma_0^* + \gamma_1^* x_1 + \gamma_2^* x_2 + \gamma_3^* x_3 \\
&= 0.15 + 0.075x_1 + 0.15x_2 + 0.05x_3 \\
P(\text{CB}|X_1 = 0, X_2 = 0, X_3 = 0) &= 0.15 \\
P(\text{CB}|X_1 = 1, X_2 = 0, X_3 = 0) &= 0.225 \\
P(\text{CB}|X_1 = 0, X_2 = 1, X_3 = 0) &= 0.3 \\
P(\text{CB}|X_1 = 0, X_2 = 0, X_3 = 1) &= 0.2 \\
P(\text{CB}|X_1 = 1, X_2 = 0, X_3 = 1) &= 0.275 \\
P(\text{CB}|X_1 = 0, X_2 = 1, X_3 = 1) &= 0.35
\end{aligned}$$

We then use these rates of outcome in a multiplicative model to determine the parameters that will lead to an equivalent data set:

$$\begin{aligned}
P(\text{CB}|X_1 = x_1, X_2 = x_2, X_3 = x_3) &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_1 x_3 + \beta_5 x_2 x_3 \\
P(\text{CB}|X_1 = 0, X_2 = 0, X_3 = 0) &= \beta_0 \\
\beta_0 &= \text{logit}(0.15) \\
&= -1.735 \\
P(\text{CB}|X_1 = 1, X_2 = 0, X_3 = 0) &= \beta_0 + \beta_1 \\
\beta_1 &= \text{logit}(0.225) - (-1.735) \\
&= 0.498 \\
P(\text{CB}|X_1 = 0, X_2 = 1, X_3 = 0) &= \beta_0 + \beta_2 \\
\beta_2 &= \text{logit}(0.3) - (-1.735) \\
&= 0.887 \\
P(\text{CB}|X_1 = 0, X_2 = 0, X_3 = 1) &= \beta_0 + \beta_3 \\
\beta_3 &= \text{logit}(0.2) - (-1.735) \\
&= 0.348 \\
P(\text{CB}|X_1 = 1, X_2 = 0, X_3 = 1) &= \beta_0 + \beta_1 + \beta_3 + \beta_4 \\
\beta_4 &= \text{logit}(0.275) - (-1.735 + 0.498 + 0.348) \\
&= -0.081 \\
P(\text{CB}|X_1 = 0, X_2 = 1, X_3 = 1) &= \beta_0 + \beta_2 + \beta_3 + \beta_5 \\
\beta_5 &= \text{logit}(0.35) - (-1.735 + 0.887 + 0.348) \\
&= -0.470
\end{aligned}$$

Thus, a logistic model with parameters $\beta_0 = -1.735$, $\beta_1 = 0.498$, $\beta_2 = 0.887$, $\beta_3 = 0.348$, $\beta_4 = -0.081$, and $\beta_5 = -0.470$ will generate the same incidence rates as an additive data set with parameters $\gamma_0^* = 0.15$, $\gamma_0^* = 0.075$, $\gamma_0^* = 0.15$, and $\gamma_0^* = 0.05$. This only holds when all interactions

between all covariates, including pairwise and higher order interactions, are included in the model. To illustrate the consequences for slightly more complicated models without inclusion of all possible interaction terms we also generated and included age for these models.

Results for data generated with sparse data for non-smoking AUP users under additive and multiplicative (OR scale) are presented in Tables 30 and 31. When linear regression is applied to truly additive data sets there is essentially no evidence of interaction between AUP use and cigarette use (all $p \geq 0.996$). Similarly, when logistic regression is applied to truly multiplicative data sets there is also no evidence of interaction (all $p > 0.97$). In both of these scenarios the incidence rate estimates for the cell corresponding to AUP users with no cigarette use are very close to the true incidence rates, with the observed differences due to rounding that occurs in order to avoid having fractions of a person in some categories. We also note that there is not statistically significant evidence of interaction for models that are of the incorrect scale (all $p > 0.54$ for linear regression on multiplicative data, all $p > 0.74$ for logistic regression on additive data). These results are very similar to what was observed for the data sets and models that did not include age. We do note a difference in the incidence rate estimates from the interaction models where scale assumptions were violated, in that these are no longer unbiased.

When another covariate is included in the model the use of an interaction model is no longer sufficient to allow for unbiased estimation of incidence rates when scale assumptions are violated.

Table 30: Example 3, Main Effect Incidence Rate Estimates in Additive, Sparse Data Sets, Age Included

Linear Regression								
			Reduced Model (Model B′)			Interaction Model (Model A′)		
True AR	True IR	LRT p-value	Used?	Est. IR	Δ	Used?	Est. IR	Δ
-0.050	0.100	0.996	x	**0.100**	**0.000**		0.103	0.003
0.000	0.150	0.986	x	**0.143**	**-0.007**		0.139	-0.011
0.050	0.200	0.996	x	**0.200**	**0.000**		0.203	0.003
0.125	0.275	0.996	x	**0.275**	**0.000**		0.278	0.003
0.200	0.350	0.997	x	**0.349**	**-0.001**		0.347	-0.003
Logistic Regression								
			Reduced Model (Model D′)			Interaction Model (Model C′)		
True AR	True IR	LRT p-value	Used?	Est. IR	Δ	Used?	Est. IR	Δ
-0.050	0.100	0.919	x	**0.127**	**0.027**		0.121	0.021
0.000	0.150	0.986	x	**0.150**	**0.000**		0.153	0.003
0.050	0.200	0.906	x	**0.191**	**-0.009**		0.200	0.000
0.125	0.275	0.746	x	**0.246**	**-0.029**		0.264	-0.011
0.200	0.350	0.748	x	**0.307**	**-0.043**		0.327	-0.023

Tables 33, 34, and 35 provide the same information as Tables 30, 31, and 32, except for non-sparse instead of sparse data sets. As was observed when age was not a factor, non-sparse data sets demonstrate greater power to detect interaction when scale assumptions are violated than sparse data sets. The smallest p-value when linear regression is applied to multiplicative data is 0.005 among the non-sparse data sets compared to 0.55 among sparse data sets, and when logistic

Table 31: Example 3, Main Effect Incidence Rate Estimates in Multiplicative (OR scale), Sparse Data Sets, Age included

Linear Regression								
			Reduced Model (Model B′)			Interaction Model (Model A′)		
True OR	True IR	LRT p-value	Used?	Est. IR	Δ	Used?	Est. IR	Δ
0.63	0.100	0.591	x	**0.047**	**-0.053**		0.069	-0.031
1.42	0.200	0.851	x	**0.217**	**0.017**		0.203	0.003
2.15	0.275	0.764	x	**0.319**	**0.044**		0.298	0.023
3.05	0.350	0.546	x	**0.399**	**0.049**		0.368	0.018
Logistic Regression								
			Reduced Model (Model D′)			Interaction Model (Model C′)		
True OR	True IR	LRT p-value	Used?	Est. IR	Δ	Used?	Est. IR	Δ
0.63	0.100	0.995	x	**0.100**	**0.000**		0.099	-0.001
1.42	0.200	0.999	x	**0.199**	**-0.001**		0.200	0.000
2.15	0.275	0.974	x	**0.276**	**0.001**		0.279	0.004
3.05	0.350	1.000	x	**0.349**	**-0.001**		0.342	-0.008

Table 32: Example 3, Main Effect Incidence Rate Estimates in Multiplicative (OR scale), Sparse Data Sets with Interaction, Age included

Linear Regression								
			Reduced Model (Model B′)			Interaction Model (Model A′)		
True OR	True IR	LRT p-value	Used?	Est. IR	Δ	Used?	Est. IR	Δ
0.63	0.100	0.751	x	**0.089**	**-0.011**		0.069	-0.031
1.00	0.150	0.273	x	**0.188**	**0.038**		0.143	-0.007
1.42	0.200	0.057	x	**0.269**	**0.069**		0.203	0.003
2.15	0.275	0.039		0.369	0.094	x	**0.298**	**0.023**
3.05	0.350	0.019		0.447	0.097	x	**0.369**	**0.019**
Logistic Regression								
			Reduced Model (Model D′)			Interaction Model (Model C′)		
True OR	True IR	LRT p-value	Used?	Est. IR	Δ	Used?	Est. IR	Δ
0.63	0.100	0.433	x	**0.122**	**0.022**		0.099	-0.001
1.00	0.150	0.470	x	**0.180**	**0.030**		0.153	0.003
1.42	0.200	0.328	x	**0.237**	**0.037**		0.200	0.000
2.15	0.275	0.337	x	**0.321**	**0.046**		0.279	0.004
3.05	0.350	0.183	x	**0.398**	**0.048**		0.342	-0.008

regression is applied to additive data it is 0.04 among non-sparse data sets compared to 0.75. Unlike the scenarios that did not include age, fitting an interaction model, even when data is not sparse, does not result in unbiased estimation of incidence rates when the model scale is misspecified.

Table 33: Example 3, Main Effect Incidence Rate Estimates in Additive, Non-Sparse Data Sets, Age Included

Linear Regression								
			Reduced Model (Model B′)			Interaction Model (Model A′)		
True AR	True IR	LRT p-value	Used?	Est. IR	Δ	Used?	Est. IR	Δ
-0.050	0.100	1.000	x	**0.101**	**0.001**		0.101	0.001
0.000	0.150	1.000	x	**0.128**	**-0.022**		0.128	-0.022
0.050	0.200	1.000	x	**0.200**	**0.000**		0.200	0.000
0.125	0.275	1.000	x	**0.275**	**0.000**		0.275	0.000
0.200	0.350	1.000	x	**0.350**	**0.000**		0.350	0.000
Logistic Regression								
			Reduced Model (Model D′)			Interaction Model (Model C′)		
True AR	True IR	LRT p-value	Used?	Est. IR	Δ	Used?	Est. IR	Δ
-0.050	0.100	0.222	x	**0.126**	**0.026**		0.122	0.022
0.000	0.150	1.000	x	**0.150**	**0.000**		0.150	0.000
0.050	0.200	0.515	x	**0.205**	**0.005**		0.208	0.008
0.125	0.275	0.0933	x	**0.270**	**-0.005**		0.276	0.001
0.200	0.350	0.0386		0.338	-0.012	x	0.346	-0.004

Table 34: Example 3, Main Effect Incidence Rate Estimates in Multiplicative (OR scale), Non-Sparse Data Sets, Age included

Linear Regression								
			Reduced Model (Model B′)			Interaction Model (Model A′)		
True OR	True IR	LRT p-value	Used?	Est. IR	Δ	Used?	Est. IR	Δ
0.63	0.100	0.0048		0.058	-0.042	x	**0.067**	**-0.033**
1.42	0.200	0.201	x	**0.193**	**-0.007**		0.188	-0.012
2.15	0.275	0.0201		0.284	0.009	x	**0.275**	**0.000**
3.05	0.350	0.0273		0.368	0.018	x	**0.359**	**0.009**
Logistic Regression								
			Reduced Model (Model D′)			Interaction Model (Model C′)		
True OR	True IR	LRT p-value	Used?	Est. IR	Δ	Used?	Est. IR	Δ
0.63	0.100	1.000	x	**0.100**	**0.000**		0.100	0.000
1.42	0.200	1.000	x	**0.200**	**0.000**		0.200	0.000
2.15	0.275	1.000	x	**0.275**	**0.000**		0.275	0.000
3.05	0.350	1.000	x	**0.350**	**0.000**		0.350	0.000

Table 35: Example 3, Main Effect Incidence Rate Estimates in Multiplicative (OR scale), Non-Sparse Data Sets with Interaction, Age included

Linear Regression								
			Reduced Model (Model B′)			Interaction Model (Model A′)		
True OR	True IR	LRT p-value	Used?	Est. IR	Δ	Used?	Est. IR	Δ
0.63	0.100	0.050		0.072	-0.028	x	**0.064**	**-0.036**
1.00	0.150	< 0.001		0.145	-0.005	x	**0.126**	**-0.024**
1.42	0.200	< 0.001		0.212	0.012	x	**0.187**	**-0.013**
2.15	0.275	< 0.001		0.305	0.030	x	**0.276**	**0.001**
3.05	0.350	< 0.001		0.389	0.039	x	**0.362**	**0.012**
Logistic Regression								
			Reduced Model (Model D′)			Interaction Model (Model C′)		
True OR	True IR	LRT p-value	Used?	Est. IR	Δ	Used?	Est. IR	Δ
0.63	0.100	< 0.001		0.110	0.010	x	**0.100**	**0.000**
1.00	0.150	< 0.001		0.162	0.012	x	**0.150**	**0.000**
1.42	0.200	< 0.001		0.213	0.013	x	**0.200**	**0.000**
2.15	0.275	< 0.001		0.290	0.015	x	**0.275**	**0.000**
3.05	0.350	< 0.001		0.366	0.016	x	**0.350**	**0.000**

Tables 36, 37, 38, and 39 provide another set of examples including age, except that the distribution of age has been set so that users of tobacco tend to be younger that abstainers. Previously ages had been distributed independent of tobacco status. In these examples there is slightly increased power to detect interaction, although not at a magnitude that would alter any of the conclusions of statistical insignificance from the previous age examples. However, the incidence rates from the regression models that include an interaction between cigarette and AUP use display more bias than was seen previously.

Table 36: Example 3, Main Effect Incidence Rate Estimates in Additive, Sparse Data Sets, Age Included, Tobacco Users Younger

Linear Regression								
			Reduced Model (Model B′)			Interaction Model (Model A′)		
True AR	True IR	LRT p-value	Used?	Est. IR	Δ	Used?	Est. IR	Δ
-0.050	0.100	0.997	x	**0.103**	**0.003**		0.101	0.001
0.000	0.150	0.971	x	**0.143**	**-0.007**		0.150	0.000
0.050	0.200	0.974	x	**0.208**	**0.008**		0.202	0.002
0.125	0.275	0.997	x	**0.278**	**0.003**		0.276	0.001
0.200	0.350	0.971	x	**0.343**	**-0.007**		0.350	0.000
Logistic Regression								
			Reduced Model (Model D′)			Interaction Model (Model C′)		
True AR	True IR	LRT p-value	Used?	Est. IR	Δ	Used?	Est. IR	Δ
-0.050	0.100	0.858	x	**0.116**	**0.016**		0.126	0.026
0.000	0.150	0.956	x	**0.150**	**0.000**		0.157	0.007
0.050	0.200	0.826	x	**0.207**	**0.007**		0.193	-0.007
0.125	0.275	0.673	x	**0.270**	**-0.005**		0.249	-0.026
0.200	0.350	0.699	x	**0.330**	**-0.020**		0.310	-0.040

Table 37: Example 3, Main Effect Incidence Rate Estimates in Multiplicative (OR scale), Sparse Data Sets, Age included, Tobacco Users Younger

Linear Regression								
			Reduced Model (Model B′)			Interaction Model (Model A′)		
True OR	True IR	LRT p-value	Used?	Est. IR	Δ	Used?	Est. IR	Δ
0.63	0.100	0.464	x	**0.086**	**-0.014**		0.057	-0.043
1.00	0.150	0.988	x	**0.146**	**0.001**		0.141	-0.009
1.42	0.200	0.829	x	**0.201**	**0.001**		0.215	0.015
2.15	0.275	0.432	x	**0.281**	**0.006**		0.314	0.039
3.05	0.350	0.433	x	**0.366**	**0.016**		0.399	0.049
Logistic Regression								
			Reduced Model (Model D′)			Interaction Model (Model C′)		
True OR	True IR	LRT p-value	Used?	Est. IR	Δ	Used?	Est. IR	Δ
0.63	0.100	0.989	x	**0.101**	**0.001**		0.104	0.004
1.00	0.150	0.993	x	**0.153**	**0.003**		0.150	0.000
1.42	0.200	0.999	x	**0.199**	**-0.001**		0.200	0.000
2.15	0.275	0.987	x	**0.270**	**-0.005**		0.274	-0.001
3.05	0.350	0.997	x	**0.349**	**-0.001**		0.351	0.001

Table 38: Example 3, Main Effect Incidence Rate Estimates in Additive, Non-Sparse Data Sets, Age Included, Tobacco Users Younger

Linear Regression								
			Reduced Model (Model B′)			Interaction Model (Model A′)		
True AR	True IR	LRT p-value	Used?	Est. IR	Δ	Used?	Est. IR	Δ
-0.050	0.100	1.000	x	**0.100**	**0.000**		0.100	0.000
0.000	0.150	1.000	x	**0.150**	**0.000**		0.150	0.000
0.050	0.200	1.000	x	**0.200**	**0.000**		0.200	0.000
0.125	0.275	1.000	x	**0.275**	**0.000**		0.275	0.000
0.200	0.350	1.000	x	**0.350**	**0.000**		0.350	0.000
Logistic Regression								
			Reduced Model (Model D′)			Interaction Model (Model C′)		
True AR	True IR	LRT p-value	Used?	Est. IR	Δ	Used?	Est. IR	Δ
-0.050	0.100	0.051	x	**0.115**	**0.015**		0.120	0.020
0.000	0.150	0.897	x	**0.159**	**0.009**		0.160	0.010
0.050	0.200	0.657	x	**0.205**	**0.005**		0.203	0.003
0.125	0.275	0.114	x	**0.275**	**0.000**		0.271	-0.004
0.200	0.350	0.0342		0.347	-0.003	x	0.341	-0.009

Table 39: Example 3, Main Effect Incidence Rate Estimates in Multiplicative (OR scale), Non-Sparse Data Sets, Age included, Tobacco Users Younger

Linear Regression								
			Reduced Model (Model B′)			Interaction Model (Model A′)		
True OR	True IR	LRT p-value	Used?	Est. IR	Δ	Used?	Est. IR	Δ
0.63	0.100	0.008		0.008	-0.020	x	**0.071**	**-0.029**
1.00	0.150	0.850	x	**0.136**	**-0.014**		0.135	-0.015
1.42	0.200	0.377	x	**0.192**	**-0.008**		0.196	-0.004
2.15	0.275	0.0195		0.275	0.000	x	**0.275**	**0.000**
3.05	0.350	0.0077		0.356	0.006	x	**0.365**	**0.015**
Logistic Regression								
			Reduced Model (Model D′)			Interaction Model (Model C′)		
True OR	True IR	LRT p-value	Used?	Est. IR	Δ	Used?	Est. IR	Δ
0.63	0.100	1.000	x	**0.100**	**0.000**		0.100	0.000
1.00	0.150	1.000	x	**0.150**	**0.000**		0.150	0.000
1.42	0.200	1.000	x	**0.200**	**0.000**		0.200	0.000
2.15	0.275	1.000	x	**0.275**	**0.000**		0.275	0.000
3.05	0.350	1.000	x	**0.350**	**0.000**		0.350	0.000

We also compared conclusions that would be drawn from several truly additive data sets to

which logistic regression was applied, and vice versa. In a sparse data set in which AUP use had the same attributable risk as moderate cigarette smoking (0.075), a logistic regression model (that did not include an interaction) would estimate the odds ratios for AUP use and moderate cigarette use to be 1.46 and 1.65, respectively. In this case, AUP use would appear to be associated with lower risk of chronic bronchitis than moderate cigarette use. However, in a data set in which all cigarette and AUP combinations occur with equal frequency, the estimated odds ratios for AUP use and moderate cigarette use are more similar (1.49 and 1.55, respectively). Among sparse data sets, if the attributable risk for AUP use is 0.1, between the true risks for moderate (0.075) and heavy cigarette use (0.15), then the estimated odds ratio for AUP use is 1.63 and is very similar to the estimated odds ratio for moderate cigarette use of 1.65. If the attributable risk for AUP use is the same as heavy cigarette use the estimated AUP odds ratio is smaller than the estimated heavy cigarette use odds ratio (2.03 vs. 2.39), and if the attributable risk for AUP use is 0.175, or even larger than that for heavy cigarette use, then logistic regression estimates the association between AUP and chronic bronchitis to be less harmful than for heavy cigarette use (2.26 vs. 2.39). These relationships are more accurately estimated when performing logistic regression on additive data sets that do not have sparsity. These results are summarized in Table 40.

Table 40: Example 3, Comparison of Interpretations When Logistic Regression Applied to Additive Data

	True AR			Est. OR		
AUP Effect Same As Moderate Smoking						
	AUP	Mod cig	Hvy cig	AUP	Mod cig	Hvy cig
Sparse	0.075	0.075	0.150	1.46	1.65	2.41
Non-sparse				1.49	1.55	2.22
AUP Effect Between Smoking Levels						
	AUP	Mod cig	Hvy cig	AUP	Mod cig	Hvy cig
Sparse	0.100	0.075	0.150	1.63	1.65	2.40
Non-sparse				1.68	1.53	2.17
AUP Effect Same As Heavy Smoking						
	AUP	Mod cig	Hvy cig	AUP	Mod cig	Hvy cig
Sparse	0.150	0.075	0.150	2.03	1.65	2.40
Non-sparse				2.10	1.49	2.11
AUP Effect Larger Than Heavy Smoking						
	AUP	Mod cig	Hvy cig	AUP	Mod cig	Hvy cig
Sparse	0.175	0.075	0.150	2.26	1.65	2.39
Non-sparse				2.33	1.48	2.09

In the last portion of this example we investigated what sample sizes would have been required to detect interaction when modeling on an incorrect scale. We investigated this in two ways; by keeping the AUP by cigarette cell counts fixed in all cells except for the AUP user/non-smoker cell, and by increasing all six cell counts proportionally. Sample sizes were increased until the LRT for interaction terms was significant at the $\alpha = 0.05$ level.

The additive data set with the most extreme effect size had an attributable risk difference for AUP use of 0.2. As seen in Table 24, row 10, when age is not included the LRT p-value under a logistic regression model for this data set is 0.281 under our sparsity conditions. As the number of AUP users who do not smoke is increased, keeping all other AUP use by cigarette use frequencies

fixed, this p-value decreases until, with 753 subjects $p = 0.048$. This is an increase of 553 subjects from the original data set. When the true attributable risk is 0.125, 2,020 non-smoking, AUP-using subjects are required to have sufficient evidence of interaction by the LRT ($p = 0.0495$). When the true attributable risk is 0.050, increasing only this cell is insufficient to detect interaction, as having 100,000 non-smoking, AUP-using subjects (in a sample that has 39,800 other subjects) only results in a LRT p-value of 0.262.

When all cell counts are increased proportionally, a total of 93,000 subjects are required to detect interaction in an additive data set with attributable risk of 0.2 (LRT $p = 0.050$). Recall that our original example included 40,000 subjects. For an attributable risk of 0.125 the required sample size in our example is 151,800 ($p = 0.049$), for attributable risk of 0.050 the required sample size is 601,600 ($p = 0.046$), and for attributable risk of -0.050 the required sample size is 296,600 ($p = 0.050$).

We also performed these power calculations for multiplicatively generated data sets in which cigarette and AUP use had an interactive association with the outcome. Recall that in these data sets the interaction was such that the odds ratio for moderate smoking among AUP users was 28% higher than among AUP non-users, and similarly, the odds ratio for heavy smoking was also set to be 28% higher among AUP users than among non-users. For the multiplicatively generated data set with the most extreme AUP main effect size (OR = 3.05), there was not statistically significant evidence of interaction (LRT $p = 0.253$) when looking at sparse data. In order to observe significant evidence of interaction, 583 non-smoking, AUP-using subjects needed to be observed with all other tobacco combinations fixed (LRT $p = 0.05$). This number is larger for interactions of lesser magnitude (1199 when odds ratios are 22% higher among AUP users) and smaller for interactions of greater magnitude (344 when odds ratios are 35% higher among AUP users).

When all cell counts were increased proportionally a total of 95,000 subjects are required to detect interaction in a multiplicative data set with AUP odds ratio of 3.05 and smoking odds ratios 28% higher among AUP users than among non-users (LRT $p = 0.049$). 145,200 (p = 0.049) for the smaller interaction of 22% higher; 63,800 (p = 0.047) for the larger interaction of 35% higher.

It is important to recall we generated data sets such that the association between age and outcome is additive when the association between tobacco use and outcome is additive. By using multiplicatively generated data sets that include age and interaction between AUP use and cigarette use we can examine the sample sizes required to detect interaction without the added complication of a covariate being modelled on the incorrect scale.

Example 4

Example 4 maintains the same framework as the basic PATH data sets from Example 3, but alters the cigarette covariate categorization that was used in regression modeling. In sparse data sets we see in Tables 41 and 42 that there is insufficient power to determine that the interaction term adds significant information to the regression models (both linear and logistic). If reduced models are implemented they tend to estimate effect sizes that imply a more beneficial association between AUP use and the outcome than exists in truth.

As an example, we will consider the case of the multiplicative data set with a true AUP odds ratio of 2.15. An analysis that dichotomized cigarette use would note some evidence of an interaction between cigarette use and AUP use (LRT $p = 0.181$), but not sufficient to conclude that the

interaction term is statistically significant. The reduced model finds that AUP use is associated with higher risk of outcome (OR = 1.12) compared to no AUP use, adjusted for cigarette use. However, we know that this is underestimating the harmful association of AUP use, since this model was set up such that the true odds ratio should be 1.42. We can also see that the interaction model is rejected due to the issue of sparse data by looking at Table 43. In non-sparse data the LRT p-values all indicate that the interaction term is highly significant in these models. Once an interaction model is fit to the data, the odds ratio estimates increase to the point that they are accurately reflecting the true relationship. This is also true for linear regression on additive models; when data is sparse there is insufficient power to conclude that the interaction is significant, but the reduced models consistently underestimate the harmful effect of AUP use.

Table 41: Example 4, OR Estimates in Multiplicative (OR scale), Sparse Data Sets with Coarse Categorization of Cigarette Use

			Reduced Model (Model B)			Interaction Model (Model A)		
True OR	True IR	LRT p-value	Used?	$\widehat{\text{OR}}$	Δ	Used?	$\widehat{\text{OR}}$	Δ
0.63	0.100	0.320	x	**0.50**	**-0.13**		0.63	0.00
1.00	0.150	0.226	x	**0.79**	**-0.21**		1.00	0.00
1.42	0.200	0.181	x	**1.12**	**-0.30**		1.42	0.00
2.15	0.275	0.128	x	**1.70**	**-0.45**		2.15	0.00
3.05	0.350	0.102	x	**2.42**	**-0.63**		3.05	0.00

Table 42: Example 4, OR Estimates in Additive, Sparse Data Sets with Coarse Categorization of Cigarette Use

			Reduced Model (Model D)			Interaction Model (Model C)		
True AR	True IR	LRT p-value	Used?	$\widehat{\text{AR}}$	Δ	Used?	$\widehat{\text{AR}}$	Δ
-0.050	0.100	0.0821	x	**-0.093**	**-0.043**		-0.050	0.000
0.000	0.150	0.0835	x	**-0.043**	**-0.043**		0.000	0.000
0.050	0.200	0.0848	x	**0.007**	**-0.043**		0.050	0.000
0.125	0.275	0.0861	x	**0.082**	**-0.043**		0.125	0.000
0.200	0.350	0.0870	x	**0.157**	**-0.043**		0.200	0.000

Table 43: Example 4, OR Estimates in Multiplicative (OR scale), Non-Sparse Data Sets with Coarse Categorization of Cigarette Use

			Reduced Model (Model B)			Interaction Model (Model A)		
True OR	True IR	LRT p-value	Used?	$\widehat{\text{OR}}$	Δ	Used?	$\widehat{\text{OR}}$	Δ
0.63	0.100	< 0.001		0.75	0.12	x	**0.63**	**0.00**
1.00	0.150	< 0.001		1.18	0.18	x	**1.00**	**0.00**
1.42	0.200	< 0.001		1.67	0.25	x	**1.42**	**0.00**
2.15	0.275	< 0.001		2.52	0.37	x	**2.15**	**0.00**
3.05	0.350	< 0.001		3.56	0.51	x	**3.05**	**0.00**

Table 44: Example 4, OR Estimates in Additive, Non-Sparse Data Sets with Coarse Categorization of Cigarette Use

			Reduced Model (Model D)			Interaction Model (Model C)		
True AR	True IR	LRT p-value	Used?	$\widehat{\text{AR}}$	Δ	Used?	$\widehat{\text{AR}}$	Δ
-0.050	0.100	< 0.001		-0.022	0.028	x	**-0.050**	**0.000**
0.000	0.150	< 0.001		0.028	0.028	x	**0.000**	**0.000**
0.050	0.200	< 0.001		0.078	0.028	x	**0.000**	**0.000**
0.125	0.275	< 0.001		0.153	0.028	x	**0.000**	**0.000**
0.200	0.350	< 0.001		0.228	0.028	x	**0.000**	**0.000**

Tables 45 and 46 show the results for sparse data sets in which we have reversed the association between AUP use and cigarette intensity, so that heavy smokers are much more likely to use AUPs than moderate smokers. The reduced models for both linear and logistic regression models are again biased, but in this scenario the bias in in the opposite direction as seen previously; all the effect size estimates overestimate the harmful association between AUP and outcome.

The amount of bias in the reduced models when using linear regression is of a consistent amount; ± 0.043 for the sparse models. This is also true for logistic models if $\hat{\beta}$ is compared to the true β coefficient, rather than looking at estimated and true odds ratios. This can be shown be focusing on the β coefficients rather than the odds ratios in Table 45.

$$
\begin{aligned}
\beta_0 &= \text{logit}(0.015) \\
&= -1.73 \\
\text{if True OR} = 0.63 \Rightarrow \beta_0 + \beta_3 &= \text{logit}(0.1) \\
\beta_3 &= \text{logit}(0.1) - (-1.73) \\
\beta_3 &= -0.463 \\
\hat{\beta}_3 &= -0.695 \\
\hat{\beta}_3 - \beta_3 &= -0.695 - (-0.463) \\
&= -0.232
\end{aligned}
$$

By performing these calculations for the other four multiplicative data sets we found that all five had $\hat{\beta}_3 - \beta_3$ between -0.231 and -0.236, which are essentially equal subject to rounding. The actual difference between $\hat{\beta}_3$ and β_3 (in the logistic case) and $\hat{\gamma}_3$ and γ_3 (in the linear case) is dependent on the frequency of AUP use by cigarette use in the population.

Table 45: Example 4, OR Estimates in Multiplicative (OR scale), Sparse Data Sets with Coarse Categorization of Cigarette Use (Reversed)

			Reduced Model (Model B)			Interaction Model (Model A)		
True OR	True IR	LRT p-value	Used?	$\widehat{\text{OR}}$	Δ	Used?	$\widehat{\text{OR}}$	Δ
0.63	0.100	0.279	x	**0.80**	**0.17**		0.63	0.00
1.00	0.150	0.209	x	**1.26**	**0.26**		1.00	0.00
1.42	0.200	0.167	x	**1.78**	**0.36**		1.42	0.00
2.15	0.275	0.127	x	**2.69**	**0.54**		2.15	0.00
3.05	0.350	0.111	x	**3.80**	**0.75**		3.05	0.00

Table 46: Example 4, OR Estimates in Additive, Sparse Data Sets with Coarse Categorization of Cigarette Use (Reversed)

			Reduced Model (Model D)			Interaction Model (Model C)		
True AR	True IR	LRT p-value	Used?	$\widehat{\text{AR}}$	Δ	Used?	$\widehat{\text{AR}}$	Δ
-0.050	0.100	0.0798	x	**-0.007**	**0.043**		-0.050	0.000
0.000	0.150	0.0810	x	**0.043**	**0.043**		0.000	0.000
0.050	0.200	0.0820	x	**0.093**	**0.043**		0.050	0.000
0.125	0.275	0.0830	x	**0.168**	**0.043**		0.125	0.000
0.200	0.350	0.0835	x	**0.243**	**0.043**		0.200	0.000

In a final extension to this example we generated data sets where cigarette use is a categorical variable taking on four values (non-smoker, light smoker, moderate smoker, and heavy smoker) but the regression models treated cigarette use as a binary variable (no/light smoking vs. moderate/heavy smoking). Results for sparse and non-sparse multiplicative (OR scale) data sets are presented in Table 47. The true odds ratios and true incidence rates used here are slightly different than in previous examples, in an effort to keep the observed incidence rates in each cell within a reasonable range.

Table 47: Example 4, OR Estimates in Multiplicative (OR scale), Sparse Data Sets with Coarse Categorization of Cigarette Use (4x2)

		Reduced Model (Model B″)			Interaction Model (Model A″)		
True OR	LRT p-value	Used?	$\widehat{\text{OR}}$	Δ	Used?	$\widehat{\text{OR}}$	Δ
0.47	< 0.001		0.50	0.03	x	**0.73**	**0.26**
1.00	< 0.001		1.09	0.09	x	**1.55**	**0.55**
1.59	< 0.001		1.76	0.17	x	**2.45**	**0.86**
2.61	< 0.001		2.96	0.35	x	**4.00**	**1.39**
3.86	< 0.001		4.47	0.61	x	**5.90**	**2.04**

Even in sparse data sets there was sufficient power to determine that the interaction terms were statistically significant for the logistic models. However, unlike previous examples, the interaction models here have larger biases than the reduced models, and for larger effect sizes massively overestimate the "true" odds ratio, while the reduced model estimates are relatively close to the truth.

Example 5

Example 5 used a different, rarer outcome than Examples 3 and 4, substituting myocardial infarction (MI) for chronic bronchitis. It also allowed tobacco use to change over time and associated different risks of MI with different durations using various types of tobacco. When cigarette use and AUP use are treated as dichotomous variables in the logistic regression model, the estimated AUP odds ratio from this logistic regression model was smaller than the true odds ratio for both 20 years of AUP use (AUP_{20}) as well as the true odds ratio for 10 years of AUP use (AUP_{10}). When this regression is performed on a non-sparse data set the estimated AUP effect falls between the effect for 10 years of use and 20 years of use (aside from the case where AUP use has no effect). It is of interest to consider these results for the most extreme of the AUP effects, when the AUP_{10} odds ratio is 6.29 and the AUP_{20} odds ratio is 8.31. These true odds ratios were set such that AUP had a more harmful association with MI than even heavy cigarette smoking. However, in the sparse data set AUP use is estimated to be less harmful than cigarette use. These estimated and true odds ratio estimates are presented in Table 48.

Table 48: Example 5, AUP Estimated and True Odds Ratios, Longitudinal Data Sets

True		Estimated			
		Sparse		Non-sparse	
AUP_{10}	AUP_{20}	AUP	Cig	AUP	Cig
1.00	1.00	0.64	4.98	1.05	4.36
1.82	2.32	1.07	4.99	2.14	4.37
3.38	4.33	1.97	4.95	3.95	4.32
4.98	6.42	2.87	4.91	5.77	4.29
6.29	8.31	3.91	4.87	7.32	4.25

We also experimented with altering the frequencies of each tobacco combination in Table 14 in order to determine whether it was possible to observe the estimated odds ratios to be larger than either of the true odds ratios for 10 years or 20 years of AUP use. A redistribution of that table

is presented in Table 49. Frequencies were changed in seven cells, with the most major changes being an increase in number of continuous concomitant heavy smokers and AUP users from 20 to 600 and an increase in the number of continuous heavy smokers who began using an AUP after 10 years on study from 10 to 600.

Table 49: Example 5, Generated Tobacco Use Over Time, Alternate Version

	Status at Year 10					
Status at Year 0	None	Mod. cig.	Hvy. cig.	AUP	Mod. cig. + AUP	Hvy. cig + AUP
None	26825	1200	0	30	0	0
Mod. cig.	1065	4160	0	50	15	0
Hvy. cig.	0	1140	4170	0	40	600
AUP	0	0	0	50	0	0
Mod. + AUP	0	0	0	10	25	0
Hvy. + AUP	0	0	0	0	20	600

We then performed logistic regression, with dichotomous cigarette use and AUP use covariates, on this data set (Table 50). Here, the estimated odds ratio for AUP use is consistently larger than either of the true odds ratios for AUP use. Examining the case where AUP use is harmful, but less harmful than all forms of cigarette use (AUP_{10} OR = 1.82, AUP_{20} OR = 2.32) we do not observe that the estimates from logistic regression reverse this relationship, as we saw previously for the most extreme AUP effect. However, in the case where AUPs and moderate smoking are set to have the same effect (AUP_{10} OR = 3.38, AUP_{20} OR = 4.33) we see that the estimated AUP odds ratio of 5.02 is slightly higher than the estimated cigarette odds ratio of 4.96.

Table 50: Example 5, AUP Estimated and True Odds Ratios, Alternative Longitudinal Data Set

True		Estimated Sparse	
AUP_{10}	AUP_{20}	AUP	Cig
1.00	1.00	1.33	4.98
1.82	2.32	2.68	4.98
3.38	4.33	5.02	4.96
4.98	6.42	7.38	4.98
6.29	8.31	9.43	4.98

Under the original distribution of cigarette and AUP statuses the logistic regression model using trichotomous categorization of cigarette use was more accurate at estimating the AUP effect than when cigarette use was dichotomized. The odds ratio estimates for AUP use fell between the true AUP odds ratios for all effect sizes except for the null effect. When there was truly no AUP effect the regression model estimated the AUP odds ratio to be 1.20, but with a 95% confidence interval from 0.64 to 2.27. Results from all five models are presented in Table 51. These estimates are consistently slightly lower than the estimates from the non-sparse data set, but not to an extreme degree.

Table 51: Example 5, AUP Estimated and True Odds Ratios, Longitudinal Data Set, Trichotomous Cigarette Categorization

True		Estimated					
		Sparse			Non-sparse		
AUP_{10}	AUP_{20}	AUP	Former Cig	Current Cig	AUP	Former Cig	Current Cig
1.00	1.00	1.20	1.25	5.33	1.10	1.32	5.32
1.82	2.32	1.98	1.33	5.34	2.25	1.33	5.39
3.38	4.33	3.75	1.26	5.33	4.19	1.33	5.41
4.98	6.42	5.48	1.28	5.30	6.18	1.33	5.43
6.29	8.31	7.71	1.21	5.30	7.90	1.33	5.42

We also attempted to redistribute frequencies of cigarette and AUP use in order to determine whether it was possible to observe estimated AUP odds ratios that were below or above the true AUP odds ratios. The frequencies shown in Table 49, which resulted in overestimated odds ratio estimates when both covariates were binary, also results in overestimated AUP odds ratios when cigarette use is trichotomous.

Table 52: Example 5, AUP Estimated and True Odds Ratios, Alternative Longitudinal Data Set, Trichotomous Cigarette Categorization

True		Estimated		
		Sparse		
AUP_{10}	AUP_{20}	AUP	Former Cig	Current Cig
1.00	1.00	1.28	1.38	5.36
1.82	2.32	2.58	1.29	5.36
3.38	4.33	4.85	1.26	5.34
4.98	6.42	7.15	1.32	5.35
6.29	8.31	9.15	1.23	5.37

These odds ratio estimates are smaller than we observed when cigarette use was dichotomized, but are still larger than even the true AUP odds ratio for 20 years of AUP use. The estimated AUP odds ratios are not as extreme as in the previous model (for the case when the AUP effect is equivalent to moderate smoking the estimated odds ratio when cigarette use is trichotomized is 4.85, compared to 5.02 previously). This, combined with the model allowing former and current smokers to have different risks, means that in the case where AUPs and moderate smoking have equivalent risks, we no longer estimate AUPs to be more harmful than cigarette use. We again observe that it possible for AUP use to be estimated to be more harmful than either of the true odds ratios for AUP use.

By altering the observed cigarette and AUP use combinations in a different way, we attempted to find a distribution that lead to consistent underestimation of the AUP odds ratio in a model with trichotomized cigarette use. This second alternative is shown in Table 53. While the AUP effect estimate was not always lower than the true effect of 10 years of AUP use, it was very near and could be observed to be consistently lower for more extreme data sets

Under this distribution the odds ratios for AUP use are estimated to be smaller than in the other two sparse data sets and in the non-sparse data set. As seen in Table 54, these estimates

Table 53: Example 5, Generated Tobacco Use Over Time, Second Alternate Version

	Status at Year 10					
Status at Year 0	None	Mod. cig.	Hvy. cig.	AUP	Mod. cig. + AUP	Hvy. cig + AUP
None	30710	1200	0	10	0	0
Mod. cig.	600	560	0	2000	15	0
Hvy. cig.	0	1140	3170	0	40	10
AUP	0	0	0	70	0	0
Mod. + AUP	0	0	0	400	55	0
Hvy. + AUP	0	0	0	0	10	10

are close to, and in some cases slightly lower than, the true odds ratio for 10 years of AUP use, whereas in the logistic regression models that used dichotomous cigarette use all the estimates fell below this lower bound.

Table 54: Example 5, AUP Estimated and True Odds Ratios, Second Alternative Longitudinal Data Set, Trichotomous Cigarette Categorization

True		Estimated		
			Sparse	
AUP_{10}	AUP_{20}	AUP	Former Cig	Current Cig
1.00	1.00	1.05	1.28	5.67
1.82	2.32	1.78	1.43	5.67
3.38	4.33	3.42	1.37	5.68
4.98	6.42	4.98	1.39	5.64
6.29	8.31	6.99	1.26	5.62

As an extension to the original sparse data set we introduced three different mortality/lost to follow-up rates among subjects who experienced MI: 5%, 10%, and 20%, with a mortality/lost to follow-up rate of 1% among subjects who were MI-free. When cigarette use was treated as a dichotomous covariate the introduction of mortality/loss to follow-up did not result in dramatic differences in the AUP odds ratio estimates. The most severe difference was in the case where the true AUP effect was more harmful than heavy cigarette smoking, and the models affected by mortality/loss to follow-up data sets estimated the AUP odds ratio to be from 5.02 - 5.06 compared to 3.91 from the initial data set (results in Appendix, Table 72). When cigarette use is treated as a trichotomous covariate the results are similar, with the introduction of mortality/loss to follow-up altering the AUP odds ratio estimates, but without dramatically changing interpretation.

4 Discussion

We began our investigation by using two examples to demonstrate that in a case-control study with two covariates it is possible to reach very different conclusions depending on whether marginal or conditional parameter estimates are used for inference. In Example 1 this was shown by fixing six of the eight possible case/cigarette/AUP status combinations such that the conditional unadjusted odds ratio for AUP use among non-smokers was 1.2 and the unadjusted marginal odds ratio for cigarette use was 1.57. We then examined different observed frequencies for controls who used both cigarettes and AUPs and for cases who used both cigarettes and AUPs. We observed that different distributions of the 140 subjects can lead to dramatically different marginal odds ratio estimates for AUP use, and that these marginal estimates do not have much in common with estimates of the conditional odds ratio. Since allowing for interaction is simply allowing the parameter estimates for a covariate to be different depending on the value of another covariate, this is not a surprising result.

While this example was simple it highlighted an important concept, that if interaction between covariates is present then the marginal parameter estimate for any of those covariates is non-informative with respect to the odds ratio estimates under the assumption of an interactive relationship. This is extremely important, as it means that the decision regarding whether to assume interaction in an analysis can or will have a powerful effect on the results and interpretation of that analysis. In Scenario 6 we would erroneously be lead to believe that AUP use was associated with reduced risk of esophageal cancer when the true relationship would be more complicated, with AUP use in actuality being associated with increased cancer risk among non-smokers.

In Example 2 we examined this phenomenon in a real world study. As illustrated in Figures 4 through 7 there are a wide variety of possible odds ratio estimates among non-smokers in the Rosenquist *et al.* study. These potential conditional odds ratios range from 0.035 to 132.1, where the minimum would represent a strong protective association between moist snuff use among non-smokers, while the maximum would indicate an extremely harmful association. Of the 1216 scenarios in which all eight combinations of case/cigarette/snuff categories were observed, 786 (64.6%) would have found statistically significant evidence of an interaction between cigarette and snuff use. This does not imply that there was a 65% chance of statistically significant interaction existing in this data set, since only one was observed (but was unknown to the reader). Nevertheless, it does illustrate that there are many possible ways in which an interaction could have been a concern in this study.

We found that conditional odds ratios can differ in somewhat counter-intuitive ways from the marginal odds ratio estimate of 0.70. If the conditional odds ratio for snuff use among non-smokers is greater than 0.70, one might assume that the conditional odds ratio for snuff use among smokers must then be less than 0.70. However, we identified data sets in which this was not the case. In one case the estimated odds ratio for an association between snuff use and OOSCC among non-smokers was 0.81 (95% CI: 0.40 - 1.65), the estimated odds ratio for the association among smokers was 9.71 (95% CI: 1.12 - 84.41), and the inclusion of interaction was statistically significant (LRT $p = 0.031$). If one considered the possibility of interaction in this data set and modeled appropriately, there would be evidence that moist snuff use was associated with significantly lower risk of OOSCC among non-smokers, while there was little evidence of an association between moist snuff use and OOSCC among smokers. The frequencies that lead to the result described above are presented in the Appendix, Table 55. We also found instances of both conditional odds ratio estimates being less than the marginal estimate, although none of these would have been statistically significant

evidence of interaction. Frequencies leading to the odds ratios of 0.67 (95% CI: 0.14 - 3.13) for snuff use among non-smokers and 0.52 (0.28 - 0.95) among smokers are in the Appendix, Table 56.

We also observed cases where both conditional odds ratios were greater than 1 while the marginal odds ratio was 0.7, and although the likelihood ratio test was not statistically significant in any of these cases with a sufficiently large sample it would be possible to have detected significant evidence of interaction. An example of this phenomenon is presented in Appendix, Table 57, and results in conditional odds ratio estimates for snuff use given non-smoking and smoking statuses of 1.21 (95% CI: 0.61 - 2.43) and 1.85 (95% CI: 0.11 - 30.00), respectively. This case is perhaps the most concerning, as interpretation of the original model would conclude that there is not sufficient evidence to conclude that moist snuff use is associated with OOSCC, or, if the marginal odds ratio had been significant, that moist snuff users have lower rates of OOSCC than non-users while analysis that includes interaction would conclude that moist snuff use is associated with higher risk of OOSCC among both smokers and non-smokers. The conclusions that would be drawn could be completely different depending on whether an interaction was taken into account, and this would have serious implications from a public health perspective. As such, we believe that it is extremely important, although not necessarily sufficient, to consider the possibility of interaction when conducting analysis from a study similar to that of cigarettes, moist snuff, and OOSCC.

Our first two examples emphasize the importance and impact of interaction assumptions when conducting statistical analysis. These assumptions are important regardless of whether sparsity is present in the data set, as a non-sparse data set that misspecifies interaction will be prone to the same biases as a sparse data set, and it is for this reason that we emphasize the need to carefully consider interaction assumptions when analyzing AUPs in the context of most observational studies. It is when one wishes to consider the possibility of interaction, but lacks statistical power to do so, that sparsity becomes a concern.

In Example 3 we addressed this concern directly. We compared data sets that had relatively few subjects who used AUP without also smoking to data sets in which there were equal numbers of the six cigarette/AUP use combinations. We focused on the impact of sparsity in relation to violations of interaction and model scale assumptions. In the basic version of this example we generated data sets such that AUP use and cigarette use were associated with risk of chronic bronchitis according to certain scale assumptions (additive relationship, multiplicative by the odds ratio scale, multiplicative by the relative risk scale) and then made inference on those data sets using different forms of regression models (linear regression, logistic regression, relative risk regression, and Poisson regression). We note that in these cases the difference between model scale and interaction is semantic; generating data such that the association between the two covariates and outcome is on the additive scale is no different than generating data such that the association is multiplicative, but with an interaction between AUP use and cigarette use that results in a data set that is identical to the additive scale.

When the scale used for the regression model did not match the scale under which data were generated we found that sparse data sets displayed reduced power to detect interaction when compared to non-sparse data sets. We also found that models that included interaction were able to estimate the incidence rates of chronic bronchitis accurately despite being on an incorrect scale, which is not surprising since, as mentioned above, a multiplicative model that accounts for interaction here is equivalent to an additive model without interaction. As seen in Tables 24 through 29, when regression models did not allow for interaction and interaction existed in the actual data, estimates of the incidence rates were biased. We also found that the magnitude of

this bias could be large enough to not only affect the incidence rate estimate, but to affect answers to more general questions such as whether AUP use was more or less harmful than cigarette use. As shown in Table 40 it is possible to have an additive, sparsely population data set in which AUP use is more harmful than moderate smoking, but when logistic regression is applied AUP use is estimated to be less harmful. The opposite was seen when we fitted linear regression to multiplicative data (Appendix Table 71), although we believe that it is relatively uncommon for researchers to choose to fit linear regression models to binary outcome data. It is also possible that given different prevalences of cigarette and AUP use combinations we could observe these relationships in the opposite directions.

While it may be tempting to conclude, based on the initial results from Example 3, that one should always include an interaction as this always led to unbiased incidence rate estimation, this is too simplistic. When we added age we found that inclusion of an interaction term in a regression model was no longer sufficient to overcome scale misspecification, and in fact in some cases led to larger bias than a model that only included main effects. This occurs because the inclusion of a third covariate in the model means that the regression model is no longer fully parameterized, unless all possible pairwise and higher dimension interactions are included. It also means that a data set on the additive scale is no longer the same as a data set on the multiplicative scale with an interaction, and hence a test for interaction will not be effective at compensating for violations of scale. Direct comparisons of regression models on different scales are not possible since neither model is nested in the other, and the methods that do exist tend to be computationally intensive and not widely implemented in existing software [18].

We saw in both cases, sparse and non-sparse, that misspecified regression models can be biased in estimation of incidence rates and effects of certain covariates, but in the case of interaction there do exist appropriate statistical tools to test for whether there is statistically significant interaction between covariates. We used the LRT, but other tests can also be used. We noticed that when we had created data sets such that there was interaction between cigarette use and AUP use on the multiplicative scale, this test was often unable to detect that interaction when sparsity was present, while it was adequately powered to do so in balanced data sets. This is another reason why we believe it is very important for researchers to present results from an analysis that includes interaction when it is reasonable to believe interaction may be present, or at least present relevant covariate combination frequencies, since in the absence of adequate power it may be necessary to evaluate evidence for interaction descriptively. This descriptive analysis can include assessing whether the parameter estimates under an interaction model differ in a scientifically meaningful way from the marginal estimates as well as checking whether any of the conclusions from a main effect-only analysis would be altered if the interaction model was used instead.

These checks should not be performed without giving due consideration for whether it is realistic to believe that interaction could exist between relevant covariates. This should include whether it is reasonable to believe that there is a biological mechanism by which interaction would occur, but even if there is no reason to believe this biological mechanism exists it can still be important to consider interaction. In Example 4 we considered cases where the categorizations of the variables used in the regression model were more coarse than the true categorizations which informed the generation of the data set. Specifically, we created the data sets with different risk for non-smokers, moderate smokers, and heavy smokers, but dichotomized cigarette use in our regression models, collapsing the "moderate" and "heavy" values into an "ever smoked" category. Additionally, we created the data sets such that the prevalence of AUP use was different for moderate smokers than for heavy smokers. When we then applied regression to these data sets, even when modeling on

the correct scale, we found that models that did not include an interaction tended to be biased, as shown in Tables 41 through 46. We found that, depending on the observed relationship between AUP and cigarette use among cigarette smokers, it was possible to under- or overestimate the true effect of AUP use when interaction was not included in the regression model. It is important to note that we created the data such that there was no interaction between cigarette use and AUP use. However, once cigarette use was collapsed, we essentially induced an interaction. Since analysis of the health effects of tobacco often involves categorizing continuous data, it is important to be aware of the possibility that interaction may be appropriate to consider even if it is not thought that there is a biological mechanism by which interaction occurs.

In this case, when cigarette use was collapsed into ever vs. never smokers, we found that including an interaction term led to unbiased estimation of the conditional odds ratio associated with AUP use among non-smokers. It is important to note, however, that the manner in which we collapsed cigarette use did not affect the categorization of cigarette use in this group, as the non-smoker group was defined the same whether or not moderate and heavy smokers were grouped together. When we generated data with four potential values for cigarette use and then collapsed heavy and moderate smokers into one group and light and non-smokers into another, the odds ratio estimates were no longer unbiased, and so including an interaction term is not necessarily sufficient to correct for categorization issues. We also note that including interaction will not have an effect on the loss of information that may come with categorization of a continuous variable.

In our final example we combined issues that we had previously addressed, such as sparsity, interaction assumptions, and categorization of continuous variables and examined them in the context of a longitudinal study. Sparsity was present due to certain tobacco use patterns being rarely observed, namely AUP use without current or past cigarette use. We fit regression models using categorizations of tobacco use that are often used in tobacco research: ever used vs. never used, and currently use vs. formerly used vs. never used. Unsurprisingly, the biases that we noticed when considering these issues individually were also present when considering them simultaneously and adding the complication of tobacco use changing over time. When dichotomizing cigarette use, using multiplicative data and modeling, and using our initial sparsity conditions, it was possible to observe estimated AUP odds ratios that were less than the true odds ratios associated with both ten and twenty years of AUP use (Table 48). Under other distributions of cigarette and AUP use over the twenty years of the study it was also possible to see the opposite case, where the estimated AUP odds ratio was greater than the true odds ratios for any length of AUP use (Table 50).

These examples demonstrate the potential impact of sparsity. We found that sparsity can reduce the power to detect interaction, potentially leading to bias if the decision to include interaction is based on statistical testing. We also found that sparsity can lead to serious bias if scale assumptions are violated. This is a major concern, both because it is very difficult to assess whether scale assumptions have been violated and because scale may not be adequately considered when dealing with binary outcomes. These concerns mean that if sparsity is likely to be present in a study it may be necessary to adapt the study design in order to perform appropriate inference.

One example of a study that opted to use a study design that we believe to be more appropriate for this topic is of the relationship between spit tobacco (snuff and chewing tobacco) use and disease mortality [9]. Here, the study population was explicitly restricted to former cigarette smokers, with the comparison of interest being those who switched from cigarettes to spit tobacco compared to those who quit entirely [9]. Sparsity would not be as much of a concern with this design since AUP use tends to be much more prevalent among former cigarette users than among those who

have never used tobacco. It is important to acknowledge the population to which this analysis is restricted, as it is a comparison of whether spit tobacco use is harmful when compared to no tobacco use among former smokers, which is not the same as assessing whether spit tobacco use is harmful. It also does not address the question of how risks associated with spit tobacco use compare to risks from cigarette use. However, we feel that the questions is addresses are relevant ones and that these restrictions in scope are necessary when encountering sparsity. Another example is a study that examined different types of cigarettes (type of tobacco, how they were manufactured, filter) were associated with lung cancer risk and restricted their subpopulation to men who used no type of tobacco other than cigarettes [2]. Other options include attempting to oversample sparse combinations, such as AUP users without a history of cigarette use, although this is dependent on, among other things, having the resources and population available to do so. In a study such as PATH, where one goal is to describe tobacco use trends and patterns, deliberate oversampling of a certain category of tobacco user might impact the ability to assess and describe those patterns, depending on the goals of the study.

We believe that this thesis demonstrates that there are serious potential issues with an AUP study that does not consider the potential implications of sparsity. Model assumptions regarding interaction and scale are difficult or impossible to verify. In the presence of sparsity, and when these assumptions are violated they can lead to misleading interpretation of results, which can then be exacerbated by the unbalanced number of observations for covariate combinations. Due to these concerns we contend that it is necessary to modify study design in order to appropriately analyze AUP use, whether by intentionally oversampling subjects for categories of use that are expected to be sparse or by restricting analysis to subpopulations where sparsity is unlikely to be present.

5 Conclusions

- Sparsity can magnify biases that occur when model assumptions, such as scale or interaction assumptions, are violated.
- Sparsity reduces power to detect interaction.
- Sparsity can affect inference to such an extent that AUP use could be found to be beneficial when it is truly harmful, or found to be less harmful than cigarette use when it is truly more harmful.
- Making modifications to the study design may be the only way to avoid the issues relating to sparsity in AUP research.

6 Future Directions and Limitations

We illustrated many issues associated with sparsity in tobacco research, but there are still directions to be investigated or investigated further. While we believe that using a design that compares AUP use to cigarette use and to no tobacco use, but only among former smokers, is an effective response to sparsity, we did not fully investigate this design. One complication associated with this approach is that consumption of many AUPs will be on a different scale than consumption of cigarettes (e.g., what quantity of smokeless tobacco should be compared to one pack per day per year of cigarette use?). Inaccuracies in attempts to determine this equivalence could have important ramifications for inference. Further research would also be valuable in determining if there are other issues with this study design and whether there are likely to be sufficient numbers of people who switch from cigarette use to the AUP in question.

If this study design is implemented and the risks of AUP use are established among switchers compared to former smokers and non-smokers, it may then be desired to estimate the risk of AUP use when it is used concomitantly with cigarettes. This scenario builds off of some topics in this thesis, such as coarse categorizations inducing interaction and the complications of measuring longitudinal tobacco use, but was not addressed in our research.

Future research could also focus on determining the number or proportion of subjects that need to be observed in a category to avoid the issues of sparsity. It might be beneficial for researchers to have a rough guideline, if possible, as to when sparsity might be a concern. We performed some rough power calculations in Example 3, but a more rigorous framework for observational studies that accounts for imbalances in covariate combinations could provide important insights. Related to these sample size issues, it could also be beneficial to better understand the advantages and disadvantages of oversampling a specific combination, as in this thesis we did not consider potential costs or other consequences of deliberate oversampling.

Our research here focused on binary outcome data and categorical covariates, but there are other settings of research in which sparsity may also be a concern. These include analysis of continuous covariates (where sparsity may be more difficult to detect), continuous outcome data (and the various possible distributions associated with that data), and survival analysis (a possible outcome when conducting a longitudinal analysis of health risks). Some of the same concerns are likely present in all of these settings, but further research could better identify and define those concerns.

This thesis focused on highlighting scenarios in which sparsity could be a concern by constructing empirical arguments rather than mathematical proofs. As such there are many scenarios which we did not consider, and future research into the theory underlying these issues could serve to provide a more comprehensive framework of the issues discussed in this thesis. Our examples were intended to illustrate scenarios in which common research approaches may prove ineffective or misleading.

7 Appendix

Table 55: Example 2, Possible Moist Snuff Use, Cigarette Use, and OOSCC Frequencies Resulting In Conditional Snuff Odds Ratio Estimates Both Greater Than Marginal Estimate

	Cases			Controls			Total
	Snuff			Snuff			
Cigarette use	Ever	Never	Total	Ever	Never	Total	
Ever	5	85	90	1	165	166	256
Never	15	26	41	64	90	154	195
Total	20	111	131	65	255	320	451

Table 56: Example 2, Possible Moist Snuff Use, Cigarette Use, and OOSCC Frequencies Resulting In Conditional Snuff Odds Ratio Estimates Both Less Than Marginal Estimate

	Cases			Controls			Total
	Snuff			Snuff			
Cigarette use	Ever	Never	Total	Ever	Never	Total	
Ever	18	72	90	54	112	166	256
Never	2	39	41	11	143	154	195
Total	20	111	131	65	255	320	451

Table 57: Example 2, Possible Moist Snuff Use, Cigarette Use, and OOSCC Frequencies Resulting In Conditional Snuff Odds Ratio Estimates Both Greater Than 1

	Cases			Controls			Total
	Snuff			Snuff			
Cigarette use	Ever	Never	Total	Ever	Never	Total	
Ever	1	89	90	1	165	166	256
Never	19	22	41	64	90	154	195
Total	20	111	131	65	255	320	451

Table 58: Example 3, Data Set 1, AUP Attributable Risk of -50

	Sparse Data Set							Non-sparse Data Set						
		Δ betw est IR and true IR							Δ betw est IR and true IR					
Regr. type	LRT sig?	NN	NY	MN	MY	HN	HY	LRT sig?	NN	NY	MN	MY	HN	HY
Add.	**No**	0	0	0	0	0	0	**No**	0	0	0	0	0	0
Mult. (OR)	**No**	-0.1	16.3	-0.7	2.4	1.7	-6.2	**Yes**	-0.1	0	0	0.1	-0.1	-0.1
Mult. (RR)	**No**	0.1	20.0	-1.2	4.2	1.8	-8.3	**Yes**	0	0.1	-0.1	0	0	0.1

Table 59: Example 3, Data Set 2, No AUP Effect

	Sparse Data Set							Non-sparse Data Set						
		Δ betw est IR and true IR							Δ betw est IR and true IR					
Regr. type	LRT sig?	NN	NY	MN	MY	HN	HY	LRT sig?	NN	NY	MN	MY	HN	HY
Add.	**No**	0	0	0	0	0	0	**No**	0	0	0	0	0	0
Mult. (OR)	**No**	-0.1	-0.1	-0.2	-0.2	-0.1	-0.1	**No**	-0.1	-0.1	0	0	-0.1	-0.1
Mult. (RR)	**No**	0	0	-0.1	-0.1	0	0	**No**	0	0	-0.1	-0.1	0	0

Table 60: Example 3, Data Set 3, AUP Attributable Risk of 50

	Sparse Data Set							Non-sparse Data Set						
		Δ betw est IR and true IR							Δ betw est IR and true IR					
Regr. type	LRT sig?	NN	NY	MN	MY	HN	HY	LRT sig?	NN	NY	MN	MY	HN	HY
Add.	**No**	0	0	0	0	0	0	**No**	0	0	0	0	0	0
Mult. (OR)	**No**	0.1	-14.1	0.3	-1.7	-1.2	5.3	**No**	5.4	-5.3	-0.6	0.5	-4.7	5.0
Mult. (RR)	**No**	0.2	-19.9	1.0	-3.8	-1.8	7.7	**Yes**	0	-0.1	-0.1	-0.3	0	-0.1

Table 61: Example 3, Data Set 4, AUP Attributable Risk of 125

	Sparse Data Set							Non-sparse Data Set						
		Δ betw est IR and true IR							Δ betw est IR and true IR					
Regr. type	LRT sig?	NN	NY	MN	MY	HN	HY	LRT sig?	NN	NY	MN	MY	HN	HY
Add.	**No**	0	0	0	0	0	0	**No**	0	0	0	0	0	0
Mult. (OR)	**No**	0.2	-31.2	0.7	-2.9	-2.7	10.6	**Yes**	-0.1	-0.1	0	0	-0.1	-0.2
Mult. (RR)	**No**	0.3	-49.9	2.4	-9.4	-5.1	16.8	**Yes**	0	0	-0.1	-0.1	0	-0.1

Table 62: Example 3, Data Set 5, AUP Attributable Risk of 200

	Sparse Data Set							Non-sparse Data Set						
		Δ betw est IR and true IR							Δ betw est IR and true IR					
Regr. type	LRT sig?	NN	NY	MN	MY	HN	HY	LRT sig?	NN	NY	MN	MY	HN	HY
Add.	**No**	0	0	0	0	0	0	**No**	0	0	0	0	0	0
Mult. (OR)	**No**	0.3	-42.7	0.7	-2.9	-3.3	14.0	**Yes**	-0.1	-0.2	0	-0.2	-0.1	-0.2
Mult. (RR)	**Yes**	0	-0.1	-0.1	-0.1	0	0.1	**Yes**	0	-0.1	-0.1	-0.1	0	0.1

Table 63: Example 3, Data Set 6, AUP OR 0.63

	Sparse Data Set							Non-sparse Data Set						
		Δ betw est IR and true IR							Δ betw est IR and true IR					
Regr. type	LRT sig?	NN	NY	MN	MY	HN	HY	LRT sig?	NN	NY	MN	MY	HN	HY
Add.	**No**	0.0	-25.0	1.0	-4.0	-3.0	9.0	**Yes**	0.0	0.0	0.0	0.0	0.0	0.0
Mult. (OR)	**No**	-0.1	0.1	-0.2	-0.4	-0.1	-0.3	**No**	-0.1	0.2	0.1	0.0	0.1	0.0
Mult. (RR)	**No**	0.0	4.8	-0.5	1.8	0.6	-3.1	**No**	-2.4	2.5	-0.8	0.7	2.4	-3.1

Table 64: Example 3, Data Set 8, AUP OR 1.42

	Sparse Data Set							Non-sparse Data Set						
		Δ betw est IR and true IR							Δ betw est IR and true IR					
Regr. type	LRT sig?	NN	NY	MN	MY	HN	HY	LRT sig?	NN	NY	MN	MY	HN	HY
Add.	**No**	0.0	20.0	0.0	3.0	2.0	-6.0	**Yes**	0.0	0.0	0.0	0.0	0.0	0.0
Mult. (OR)	**No**	-0.1	0.2	0.0	-0.3	-0.1	0.0	**No**	-0.2	0.0	0.1	-0.1	-0.1	0.0
Mult. (RR)	**No**	0.0	-8.1	0.8	-3.2	-1.5	3.7	**No**	3.0	-3.1	1.0	-1.2	-3.9	3.0

Table 65: Example 3, Data Set 9, AUP OR 2.15

	Sparse Data Set							Non-sparse Data Set						
		Δ betw est IR and true IR							Δ betw est IR and true IR					
Regr. type	LRT sig?	NN	NY	MN	MY	HN	HY	LRT sig?	NN	NY	MN	MY	HN	HY
Add.	**No**	0.0	39.0	-1.0	4.0	3.0	-13.0	**Yes**	0.0	0.0	0.0	0.0	0.0	0.0
Mult. (OR)	**No**	-0.1	0.1	-0.2	0.1	0.1	-0.2	**No**	-0.1	0.1	-0.2	0.1	0.3	0.0
Mult. (RR)	**No**	0.2	-26.4	2.2	-7.9	-3.6	10.7	**Yes**	0.0	0.0	-0.1	-0.1	0.0	-0.2

Table 66: Example 3, Data Set 10, AUP OR 3.05

	Sparse Data Set							Non-sparse Data Set						
		Δ betw est IR and true IR							Δ betw est IR and true IR					
Regr. type	LRT sig?	NN	NY	MN	MY	HN	HY	LRT sig?	NN	NY	MN	MY	HN	HY
Add.	**No**	-0.1	0.0	-0.2	0.0	-0.1	0.3	**Yes**	0.0	0.0	0.0	0.0	0.0	0.0
Mult. (OR)	**No**	-0.1	0.5	-0.2	0.0	-0.1	0.3	**No**	-0.1	0.0	0.1	0.0	0.1	-0.2
Mult. (RR)	**No**	0.3	-50.0	4.0	-13.0	-7.1	17.5	**Yes**	0.0	-0.1	-0.1	-0.5	0.0	-0.3

Table 67: Example 3, Data Set 11, AUP RR 0.67

	Sparse Data Set							Non-sparse Data Set						
		Δ betw est IR and true IR							Δ betw est IR and true IR					
Regr. type	LRT sig?	NN	NY	MN	MY	HN	HY	LRT sig?	NN	NY	MN	MY	HN	HY
Add.	**No**	0.0	-31.0	1.0	-6.0	-4.0	14.0	**Yes**	0.0	0.0	0.0	0.0	0.0	0.0
Mult. (OR)	**No**	0.1	-4.0	0.5	-1.9	-0.8	3.4	**No**	2.0	-2.0	0.8	-0.8	-3.1	2.8
Mult. (RR)	**No**	0.0	0.7	-0.1	-0.1	0.0	0.3	**No**	-0.3	0.3	0.1	-0.2	0.0	-0.1

Table 68: Example 3, Data Set 13, AUP RR 1.33

	Sparse Data Set							Non-sparse Data Set						
		Δ betw est IR and true IR							Δ betw est IR and true IR					
Regr. type	LRT sig?	NN	NY	MN	MY	HN	HY	LRT sig?	NN	NY	MN	MY	HN	HY
Add.	**No**	0.0	31.0	-1.0	6.0	4.0	-14	**Yes**	0.0	0.0	0.0	0.0	0.0	0.0
Mult. (OR)	**No**	-0.1	9.3	-0.9	3.4	1.3	-6.1	**No**	-3.5	3.4	-1.3	1.1	4.5	-4.6
Mult. (RR)	**No**	0.0	-0.5	-0.1	0.1	0.0	-0.1	**No**	0.2	-0.3	-0.1	0.1	0.0	-0.1

Table 69: Example 3, Data Set 14, AUP RR 1.83

	Sparse Data Set							Non-sparse Data Set						
		Δ betw est IR and true IR							Δ betw est IR and true IR					
Regr. type	LRT sig?	NN	NY	MN	MY	HN	HY	LRT sig?	NN	NY	MN	MY	HN	HY
Add.	**Yes**	0.0	0.0	0.0	1.0	0.0	0.0	**Yes**	0.0	0.0	0.0	0.0	0.0	0.0
Mult. (OR)	**No**	-0.2	36.3	-2.6	11.1	4.9	-19.5	**Yes**	-0.1	-0.1	0.0	-0.1	-0.1	-0.2
Mult. (RR)	**No**	0.0	-0.3	-0.1	-0.1	0.0	0.4	**No**	0.2	0.0	0.1	0.3	0.0	0.4

Table 70: Example 3, Data Set 15, AUP RR 2.33

	Sparse Data Set							Non-sparse Data Set						
		Δ betw est IR and true IR							Δ betw est IR and true IR					
Regr. type	LRT sig?	NN	NY	MN	MY	HN	HY	LRT sig?	NN	NY	MN	MY	HN	HY
Add.	**Yes**	0.0	0.0	0.0	0.0	0.0	0.0	**Yes**	0.0	0.0	0.0	0.0	0.0	0.0
Mult. (OR)	**Yes**	-0.1	0.0	-0.2	-0.3	-0.1	-0.3	**Yes**	-0.1	-0.2	0.0	0.0	-0.1	0.3
Mult. (RR)	**No**	0.0	-0.4	-0.1	0.1	0.0	0.1	**No**	0.2	-0.1	-0.1	0.1	0.0	0.1

Table 71: Example 3: Comparison of Interpretations When Linear Regression Applied to Multiplicative Data

	True OR			Est. AR		
AUP Effect Less Harmful Than Moderate Smoking						
	AUP	Mod cig	Hvy cig	AUP	Mod cig	Hvy cig
Sparse	1.45	1.64	2.43	0.074	0.074	0.074
Non-sparse				0.069	0.083	0.164
AUP Effect Same As Moderate Smoking						
	AUP	Mod cig	Hvy cig	AUP	Mod cig	Hvy cig
Sparse	1.64	1.64	2.43	0.101	0.074	0.152
Non-sparse				0.095	0.086	0.169
AUP Effect Between Smoking Levels						
	AUP	Mod cig	Hvy cig	AUP	Mod cig	Hvy cig
Sparse	2.20	1.64	2.43	0.169	0.074	0.153
Non-sparse				0.160	0.093	0.177
AUP Effect Same As Heavy Smoking						
	AUP	Mod cig	Hvy cig	AUP	Mod cig	Hvy cig
Sparse	2.43	1.64	2.43	0.193	0.074	0.154
Non-sparse				0.183	0.094	0.180

Table 72: Example 5, AUP Estimated and True Odds Ratios, Under Dichotomous Cigarette Categorization with Mortality

True		Estimated AUP			
AUP_{10}	AUP_{20}	No Mortality	1% vs. 5%	1% vs. 10%	1% vs. 20%
1.00	1.00	0.64	0.61	0.58	0.51
1.82	2.32	1.07	1.12	1.18	1.04
3.38	4.33	1.97	2.08	2.06	2.09
4.98	6.42	2.87	2.82	2.90	3.04
6.29	8.31	3.91	5.02	5.06	5.02

Table 73: Example 5, AUP Estimated and True Odds Ratios, Under Trichotomous Cigarette Categorization with Mortality

True		Estimated AUP			
AUP_{10}	AUP_{20}	No Mortality	1% vs. 5%	1% vs. 10%	1% vs. 20%
1.00	1.00	1.20	1.12	1.07	0.92
1.82	2.32	1.98	2.06	2.20	1.96
3.38	4.33	3.75	3.93	3.89	3.96
4.98	6.42	5.48	5.37	5.64	5.91
6.29	8.31	7.71	6.52	6.60	6.55

Table 74: Sparse Data Set Resulting in AR=0.05

	CB^+			CB^-			Total
	AUP			AUP			
Cigarette status	Yes	No	Total	Yes	No	Total	
Heavy	280	990	1270	520	2310	2830	4100
Moderate	220	742	962	580	2558	3138	4100
None	40	4740	4790	160	26860	27020	31800
Total	540	6472	7012	1260	31728	32988	40000

Table 75: Sparse Data Set Resulting in OR=1.42

	CB^+			CB^-			Total
	AUP			AUP			
Cigarette status	Yes	No	Total	Yes	No	Total	
Heavy	302	990	1292	498	2310	2808	4100
Moderate	234	742	976	566	2558	3124	4100
None	40	4740	4780	160	26860	27020	31800
Total	576	6472	7048	1224	31728	32952	40000

Table 76: Sparse Data Set Resulting in RR=1.33

	CB^+			CB^-			Total
	AUP			AUP			
Cigarette status	Yes	No	Total	Yes	No	Total	
Heavy	319	990	1309	481	2310	2791	4100
Moderate	239	742	981	561	2558	3119	4100
None	40	4740	4780	160	26860	27020	31800
Total	598	6472	7070			32930	40000

Table 77: Non-sparse Data Set Resulting in AR=0.05

	CB^+			CB^-			Total
	AUP			AUP			
Cigarette status	Yes	No	Total	Yes	No	Total	
Heavy	2333	2000	2333	4334	4667	9001	13334
Moderate	1833	1500	3333	4834	5167	10001	13334
None	1333	1000	4333	5333	5666	10999	13332
Total	5499	4500	9999	14501	15500	30001	40000

Table 78: Non-sparse Data Set Resulting in OR=1.42

	CB^+			CB^-			Total
	AUP			AUP			
Cigarette status	Yes	No	Total	Yes	No	Total	
Heavy	2520	2000	4520	4147	4667	8814	13334
Moderate	1947	1500	3447	4720	5167	9887	13334
None	1333	1000	2333	5333	5666	10999	13332
Total	5800	4500	10300	14200	15500	29700	40000

Table 79: Non-sparse Data Set Resulting in RR=1.33

	CB^+			CB^-			Total
	AUP			AUP			
Cigarette status	Yes	No	Total	Yes	No	Total	
Heavy	2660	2000	4660	4007	4667	8674	13334
Moderate	1993	1500	3493	4674	5167	9841	13334
None	1333	1000	2333	5333	5666	10999	13332
Total	5986	4500	10486	14014	15500	29514	40000

References

[1] Family Smoking Prevention and Tobacco Control Act, Pub. L. No. 111-31, HR 1256, 2009.

[2] E Benhamou, S Benhamou, A Auquier, and R Flamant. Changes in patterns of cigarette smoking and lung cancer risk: results of a case-control study. *British journal of cancer*, 60(4):601, 1989.

[3] CDC. Use of smokeless tobacco among adults — United States, 1991. *MMWR*, 42:263–266, 1993.

[4] CDC. Cigarette smoking among adults - United States, 2002. *MMWR*, 53(20):427–437, 2004.

[5] CDC. Cigarette smoking among adults — United States, 2006. *MMWR*, 56:1157–1161, 2007.

[6] Rudolf J Freund. The case of the missing cell. *The American Statistician*, 34(2):94–98, 1980.

[7] Yolanda I Garces, Ping Yang, Julia Parkinson, Xinghua Zhao, Jason A Wampfler, Jon O Ebbert, and Jeff A Sloan. The relationship between cigarette smoking and quality of life after lung cancer diagnosis. *CHEST Journal*, 126(6):1733–1741, 2004.

[8] Manal M Hassan, James L Abbruzzese, Melissa L Bondy, Robert A Wolff, Jean-Nicolas Vauthey, Peter W Pisters, Douglas B Evans, Rabia Khan, Renato Lenzi, Li Jiao, and Donghui Li. Passive smoking and the use of noncigarette tobacco products in association with risk for pancreatic cancer. *Cancer*, 109(12):2547–2556, 2007.

[9] S Jane Henley, Cari J Connell, Patricia Richter, Corinne Husten, Terry Pechacek, Eugenia E Calle, and Michael J Thun. Tobacco-related disease mortality among men who switched from cigarettes to spit tobacco. *Tobacco control*, 16(1):22–28, 2007.

[10] Harry A Lando, C Keith Haddock, Robert C Klesges, G Wayne Talcott, and Joni Jensen. Smokeless tobacco use in a population of young adults. *Addictive behaviors*, 24(3):431–437, 1999.

[11] George A Milliken and Dallas E Johnson. *Analysis of Messy Data*, volume 1. Lifetime Learning Publications, 1984.

[12] National Institute of Health. Federal register — proposed collection; comment request; population assessment of tobacco and health (PATH) study, 2012. `https://www.federalregister.gov/articles/2012/05/18/2012-12017/proposed-collection-comment-request-population-assessment-of-tobacco-and-health-path-study`, 10/26/2012.

[13] Federal Business Opportunities. Family smoking preventation and tobacco control act national longitudinal study, 2011. `https://www.fbo.gov/index?s=opportunity&mode=form&id=2a09a499283c9c18d4c66b3ea67a64ef&tab=core&_cview=1`,10/26/2012.

[14] Kerstin Rosenquist, Johan Wennerberg, Elsy-Britt Schildt, Anna Bladström, Bengt Göran Hansson, and Gunilla Andersson. Use of swedish moist snuff, smoking and alcohol consumption in the aetiology of oral and oropharyngeal squamous cell carcinoma. a population-based case-control study in southern sweden. *Acta oto-laryngologica*, 125(9):991–998, 2005.

[15] JC Schroeder, AF Olshan, R Baric, GA Dent, CR Weinberg, B Yount, JR Cerhan, CF Lynch, LM Schuman, PE Tolbert, et al. A case–control study of tobacco use and other non-occupational risk factors for t (14; 18) subtypes of non-hodgkin's lymphoma (united states). *Cancer Causes and Control*, 13(2):159–168, 2002.

[16] SR Searle, FM Speed, and GA Milliken. Population marginal means in the linear model: an alternative to least squares means. *The American Statistician*, 34(4):216–221, 1980.

[17] Stephen Soldz, Dana Joy Huyser, and Elizabeth Dorsey. Characteristics of users of cigars, bidis, and kreteks and the relationship to cigarette use. *Preventive medicine*, 37(3):250–258, 2003.

[18] Gerald van Belle. *Statistical Rules of Thumb*. Wiley Series in Probability and Statistics, second edition, 2008.

[19] W Douglas Weaver, Leonard A Cobb, Alfred P Hallstrom, Carol Fahrenbruch, Michael K Copass, and Roberta Ray. Factors influencing survival after out-of-hospital cardiac arrest. *Journal of the American College of Cardiology*, 7(4):752–757, 1986.

[20] Westat. Westat - behavioral health research - policy support, 2012. `http://www.westat.com/Westat/expertise/health_and_medical/behavioral_health/policy.cfm`, 10/26/2012.

[21] David W Wetter, Jennifer B McClure, Carl de Moor, Ludmila Cofta-Gunn, Scott Cummings, Paul M Cinciripini, and Ellen R Gritz. Concomitant use of cigarettes and smokeless tobacco: prevalence, correlates, and predictors of tobacco cessation. *Preventive medicine*, 34(6):638–648, 2002.

[22] Jian-Min Yuan, J Esteban Castelao, Manuela Gago-Dominguez, Mirni C Yu, and Ronald K Ross. Tobacco use in relation to renal cell carcinoma. *Cancer Epidemiology Biomarkers & Prevention*, 7(5):429–433, 1998.

[23] Tongzhang Zheng, Kenneth P Cantor, Yawei Zhang, Brian CH Chiu, and Charles F Lynch. Risk of brain glioma not associated with cigarette smoking or use of other tobacco products in Iowa. *Cancer Epidemiology Biomarkers & Prevention*, 10(4):413–414, 2001.