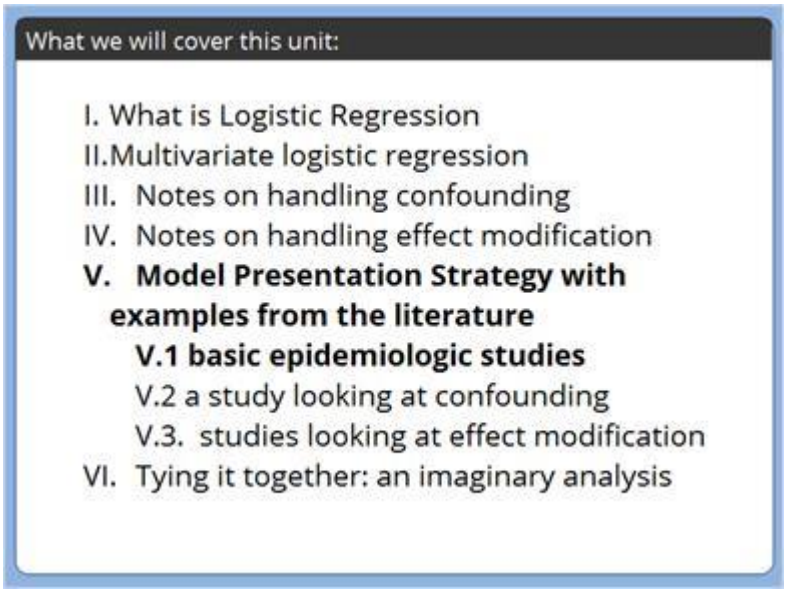
Logistic Regression - Part 2

1.1 Logistic Regression

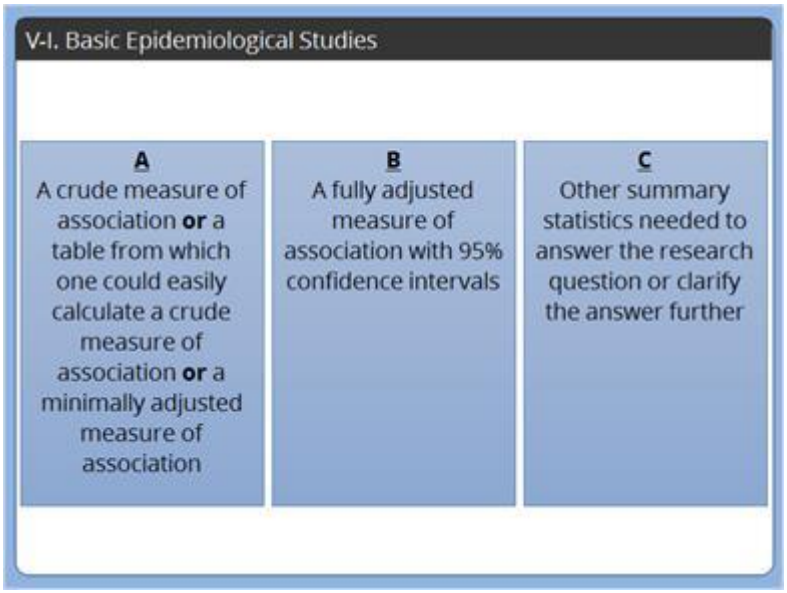
We continue our series on logistic regression by now reviewing model presentation strategies that incorporate the topics of confounding and effect modification.

1.2 What we will cover this unit:



In the game of bridge, a saying goes that "Every hand is different." It's the same for data analysis. There is an art involved in summarizing data well, and the summary may depend quite a lot on the questions being posed and the data itself.

1.3 V-I. Basic Epidemiological Studies



If our hypothesis involves the association of one primary exposure and one primary outcome, all manuscripts should minimally contain components A and B below. Most will contain C.

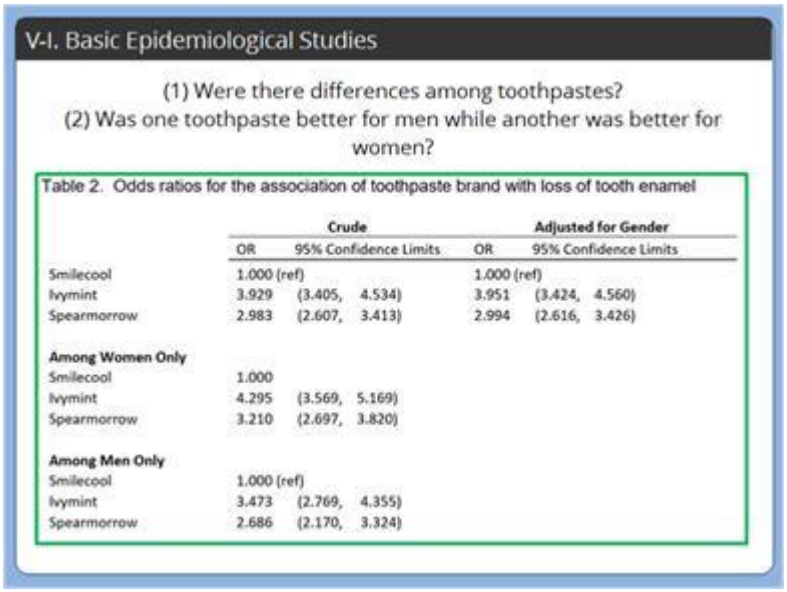
A. A crude measure of association **or** a table from which one could easily calculate a crude measure of association **or** a minimally adjusted measure of association eg: age-adjusted measure of association. Transparency, however, should always be optimized. The reader should be able to see clearly what you have. For example if both the exposure and outcome are dichotomous then, the reader should be able to easily tell the number of people in each cell of the 2x2 table, even if it's not presented as a 2x2 table. In a cohort study, the reader should easily be able to tell the number of events and the person-years of follow-up in each exposure group.

and

B. A fully adjusted measure of association with 95% confidence intervals.

C. Other summary statistics needed to answer the research question or clarify the answer further.

1.4 V-I. Basic Epidemiological Studies

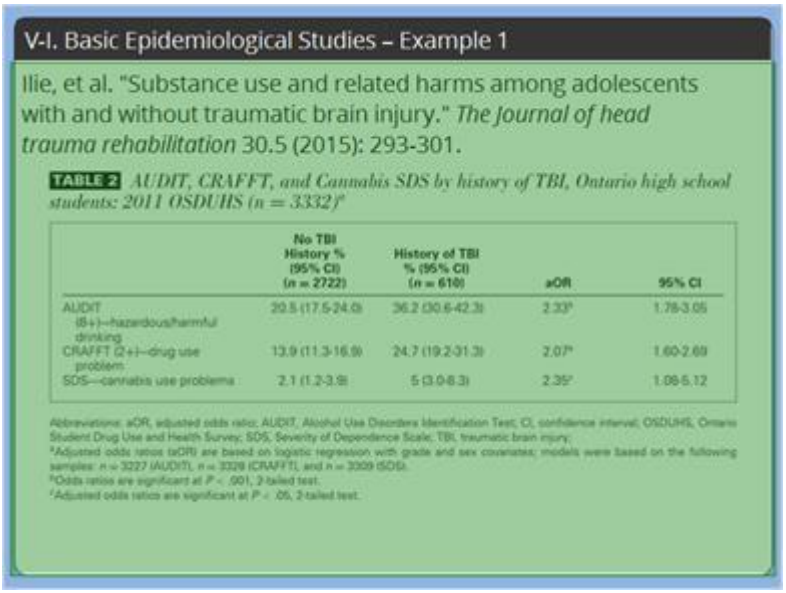


If we go back to our toothpaste example, we wanted to know if (1) there were differences among toothpastes and (2) if one toothpaste was better for men while another was better for women. We decide to present the data as follows: (Note we have to run main effects models without interaction terms. These were not shown earlier. Also we assume that all relevant counts were already given in Table 1).

This table 2 presents the essential components - the crude odds ratios and the adjusted odds ratio

Let's now look at some real examples from the literature. In each of the following examples, note the research question being asked and exactly how the authors chose to present their primary results. Please read everything, including and especially the footnotes on the tables.

1.5 V-I. Basic Epidemiological Studies – Example 1



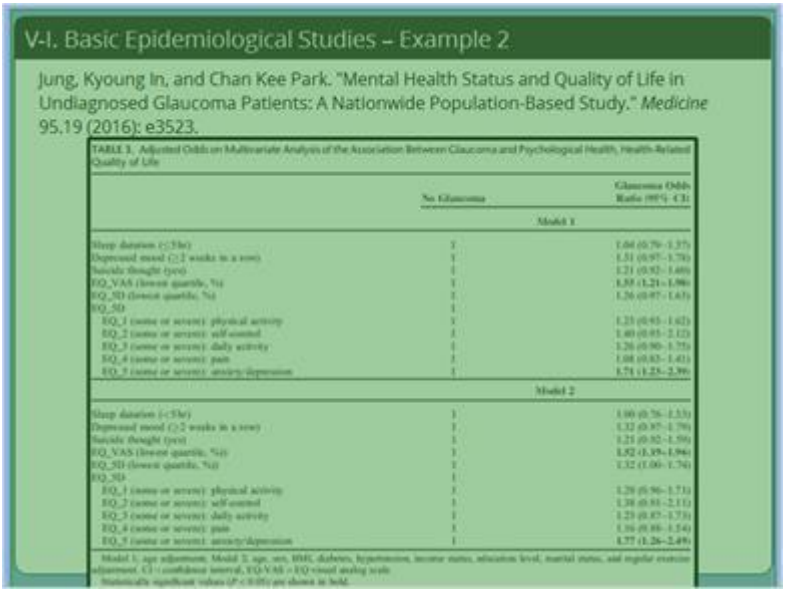
We start by looking at an example from the literature of a basic epidemiological study. The paper is entitled substance use and related harms among adolescents with and without traumatic brain injury. The objective was to look at the relationship between self-reported lifetime traumatic brain injury (TBI) and drug and alcohol use and

associated harms, using an epidemiological sample of Canadian adolescents.

Click on the slide or click the resources table to view the PDF of this article. We highlighted the area that responds to Table 2. What do we see in this table?

The prevalence and adjusted odds of substance use according to TBI status. So we are able to calculate the crude OR if we wanted, and then we have the adjusted model presented.

1.6 V-I. Basic Epidemiological Studies – Example 2



Now let's look at a 2nd example. This comes from an article titled, Mental health status and quality of life in undiagnosed glaucoma patients: a nationwide population based study.

Click on the slide or in the resource tab to get the article by Jung et al.

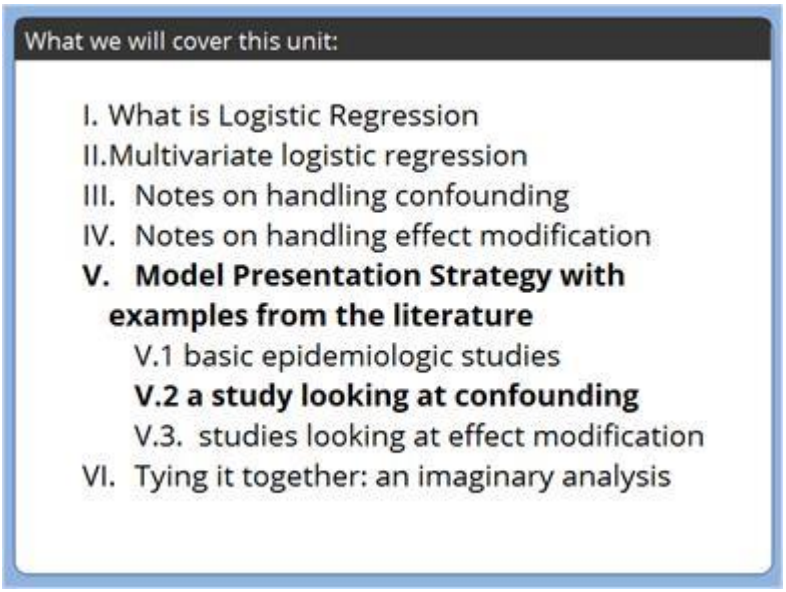
So the objective of this study was to investigate the association between mental health status or QoL and undiagnosed glaucoma, along with the effects of visual acuity or visual field damage.

The area of the results we want to highlight come from Table 3, which presents the differences in psychological health and QoL between the subjects with and without glaucoma after adjusting for age (Model 1) and age, sex, body mass index, diabetes, hypertension, income status, education level, marital status, and regular exercise (Model 2: demographic factors with P value <0.05 in univariate analyses). In both Models 1 and 2, glaucoma subjects were more likely than those without glaucoma to have some or severe problems with anxiety/depression (Model 1: OR, 1.71; 95% CI 1.23-2.39, Model 2: OR 1.77; 95% CI, 1.26-2.49)

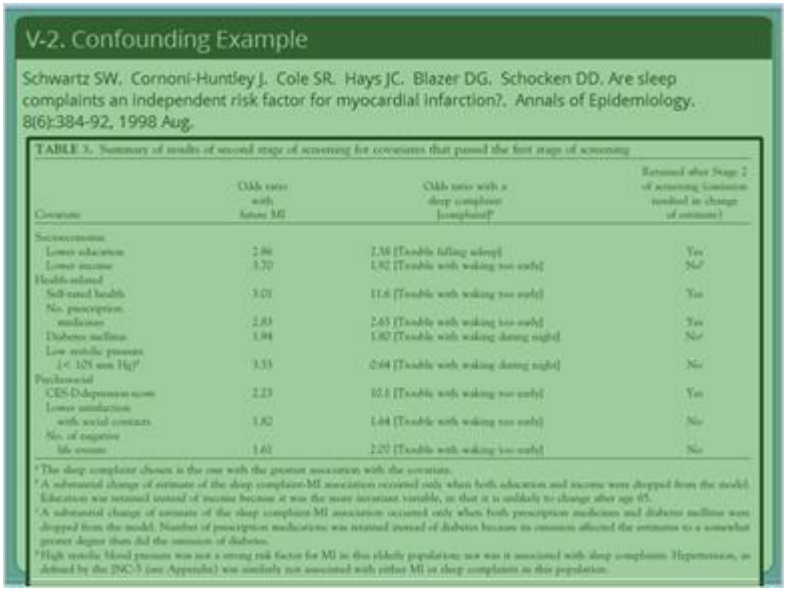
As we look at this table, what are the 1's under the no glaucoma column?

	A. The author was probably using them to denote that the odds ratios for glaucoma vs no glaucoma (i.e. no glaucoma is the reference group), but the column is probably unnecessary.

1.7 What we will cover this unit:

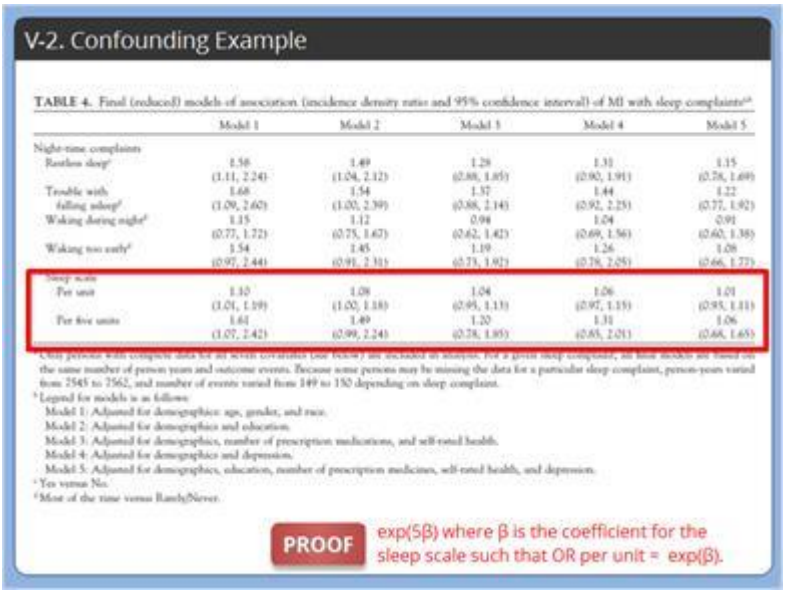
	Now let's look at an example of a study looking at confounding.

1.8 V-2. Confounding Example

	*Now let's look at an example of investigating and presenting confounding in an article. This comes from an article titled, Are sleep complaints an independent risk factor for myocardial infarction.* *Click on the slide or in the resource tab to get the article by Schwartz et al. We highlighted the areas in yellow that are important for you to read.* **The stated objective to this study was t**o investigate whether subjective sleep complaints are an independent predictor of myocardial infarction (MI) in a community of older adults *and to gain clues as to why the association between sleep complaints and incident MI exists.* **From the Methods**: Nineteen variables were evaluated to determine if they could explain the sleep-complaint-MI association. , and these were grouped into three larger groups (socioeconomic, health related, and psychosocial). In addition to a basic model adjusting for age, gender, and race (Model 1), we then defined three models,

	corresponding to our covariate groups, that is, a socioeconomic model (Model 2), a health-related model (Model 3), and a psychosocial model (Model 4). Age, gender, and race were not subjected to screening and were included in all models. Covariates were screened in two stages, which you can read in detail in the article. **From the Results**: Results of our screening for confounders are presented in Table 3. Nine variables met the criterion of being associated with MI and at least one sleep question. …Of the nine potential confounders, four (education, depression, number of prescription medicines, and self rated health) remained confounders after stage II screening.

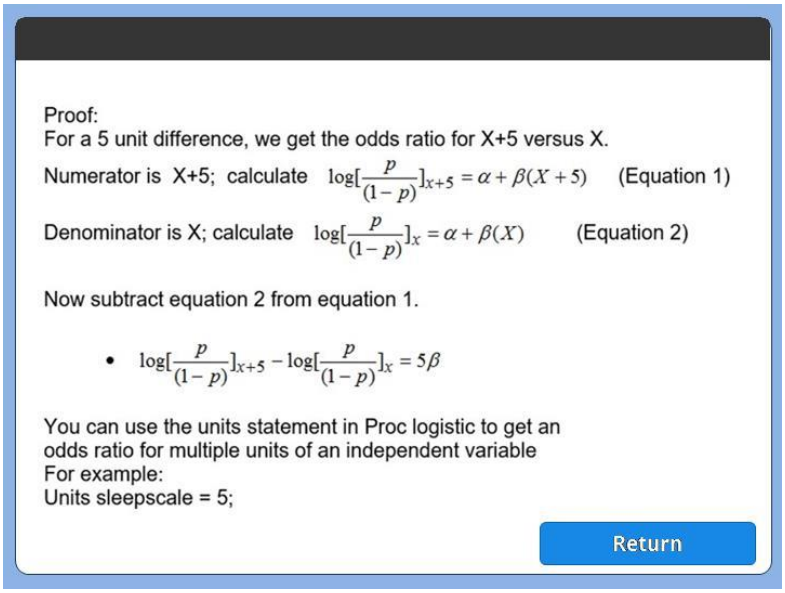
1.9 V-2. Confounding Example



To allow for direct comparison among the IDRs, only persons with complete data for the four covariates were included in the final models (Table 4).

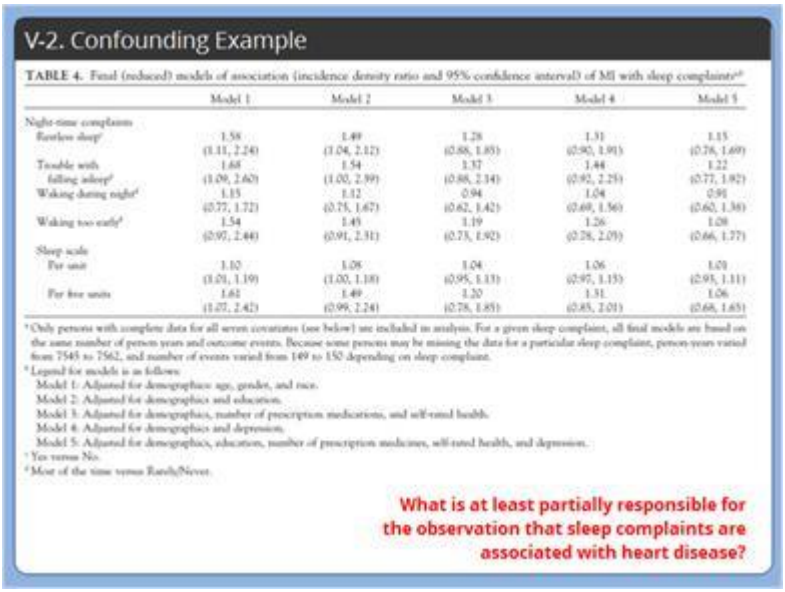
*1. In Table 4, odds ratios are given per unit of a sleep study and per five units of the sleep scale. How is this done?

Answer: the general formula for the odds ratio for a 5 unit change is simply exp(5β) where β is the coefficient for the sleep scale such that OR per unit = exp(β).



Proof

1.10 V-2. Confounding Example



2. Based on table 4 above, what is at least partially responsible for the observation that sleep complaints are associated with heart disease?

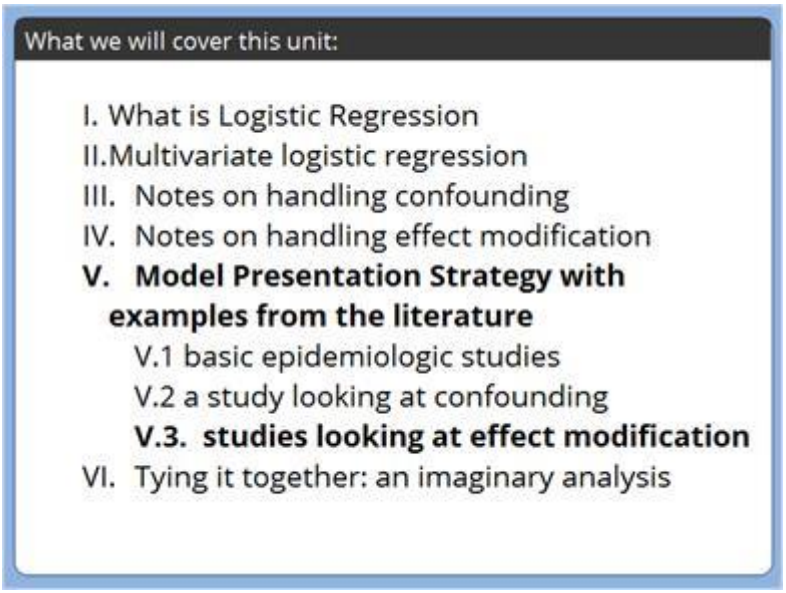
Multiple choice

A. Physical health

B. Mental health

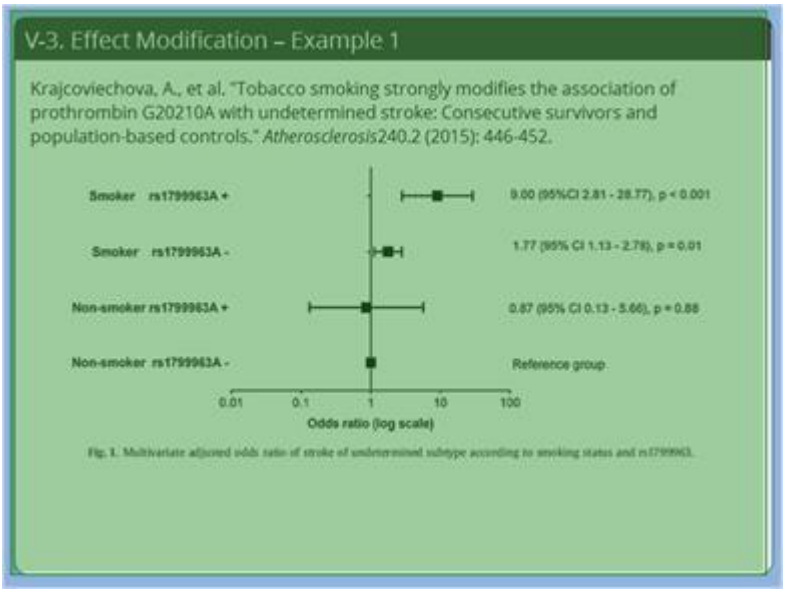
*C. Both physical and mental health. Compare Model 1 to Models 3, 4 and 5.

1.11 What we will cover this unit:



Now let's look at a couple examples of effect modification with a logistic regression study.

1.12 V-3. Effect Modification – Example 1



Let's now look at a simple example to examine effect modification.

Please click on the screen or the resource tab to get the article. You will want to read some of the background we've highlighted in the paper so you have some context for the purpose of this study.

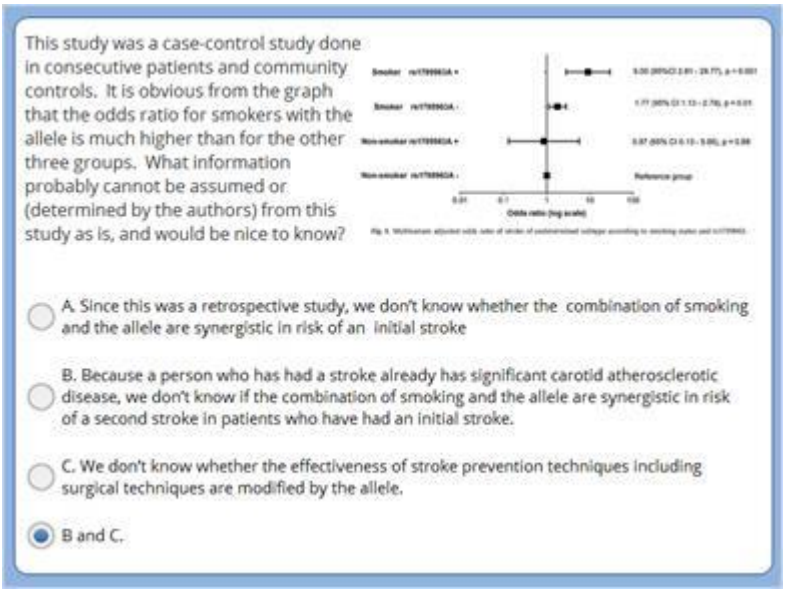
For the objective the authors state, "We evaluated the differences in the distribution of rs6025 and rs1799963 polymorphisms according to ischemic stroke subtypes and their interaction with smoking."

Again, I suggest reading some background so you know what's going on.

From the Methods: …interaction terms between smoking and carriage of at least one copy of the respective A allele (rs1799963, rs6025; regardless of homozygosity or heterozygosity) were tested in multivariate adjusted logistic regression. The odds ratio of undetermined stroke was calculated according to smoking status and carriage of the rs1799963A allele, using non-smokers non-carriers as reference category, and adjusting for other cardiovascular risk factors.

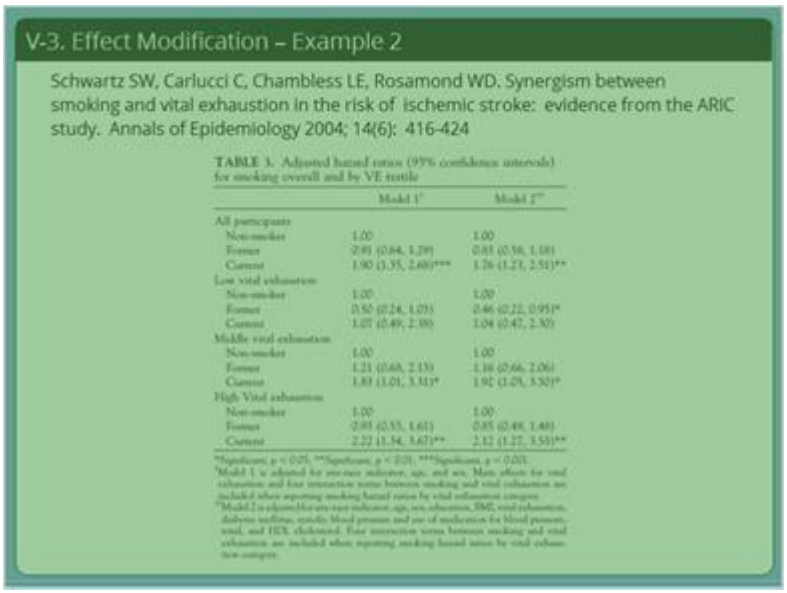
Figure 1 shows the multivariate adjusted odds ratio of stroke according to smoking status and subtype.

1.13 This study was a case-control study done in consecutive patients and community controls. It is obvious from the graph that the odds ratio for smokers with the allele is much higher than for the other three groups. What information probably cannot be assumed or (determined by the authors) from this study as is, and would be nice to know?



This study was a case-control study done in consecutive patients and community controls. It is obvious from the graph that the odds ratio for smokers with the allele is much higher than for the other three groups. What information probably cannot be assumed or (determined by the authors) from this study as is, and would be nice to know?

1.14 V-3. Effect Modification – Example 2



On to our next example. Again, click on the slide or click on the resources tab to access the file labeled Schwartz 2. This is our second effect modification example.

Note that in this study Hazard ratios are similar in meaning to incidence density ratios. They are computed using Proc PHREG in SAS. However once you have the output, you can work it exactly the way that you would work with the logistic output. The only difference is that exp(β) =hazard ratio instead of the odds ratio.

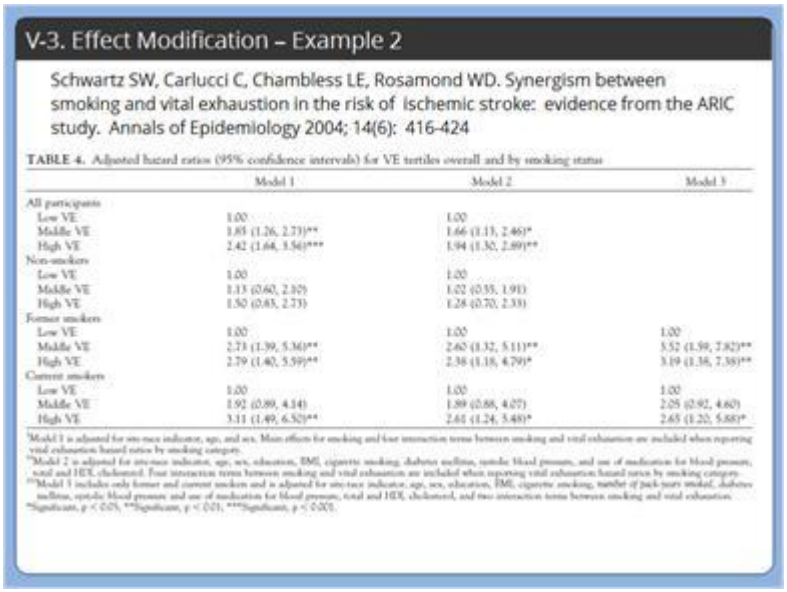
The objective of this study was to examine the synergism between vital exhaustion and cigarette smoking in producing ischemic stroke.

Vital exhaustion (VE), a state characterized by unusual fatigue, irritability, and feelings of demoralization, was measured by a 21-item inventory of symptoms.

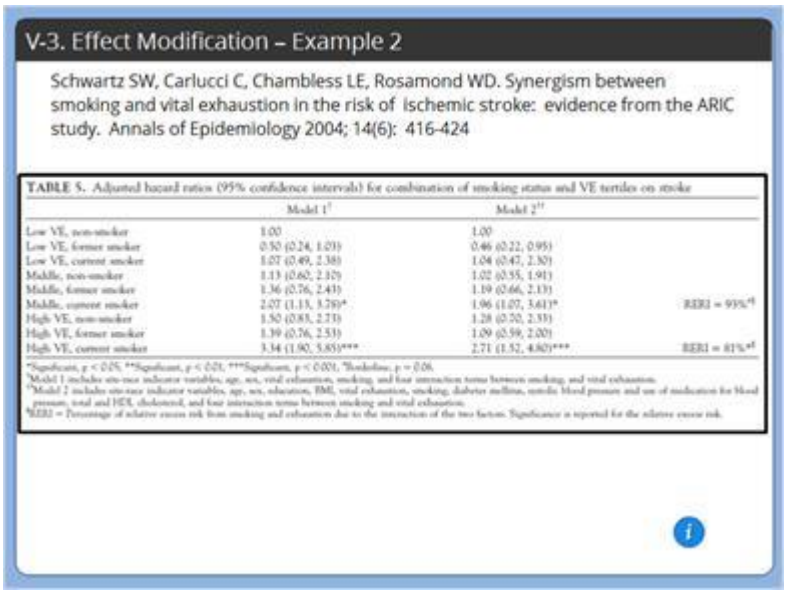
Table 3 shows us smoking-related hazard ratios overall and by category of vital exhaustion. Models 1

	and 2 included main effects of VE and smoking categories and four interaction terms created by crossing VE tertile with smoking status category.

1.15 V-3. Effect Modification – Example 2

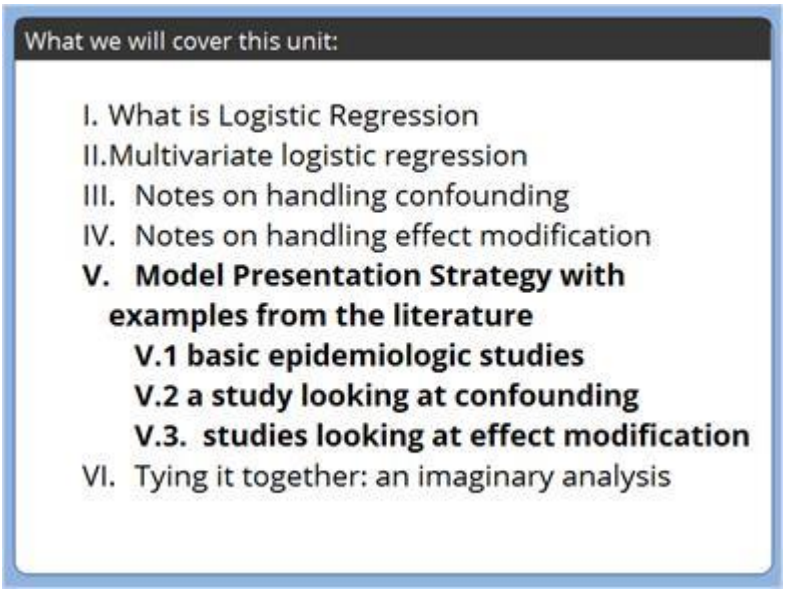
	Model 3 repeated Model 2 for current and former smokers only and also adjusted for pack-years. It was run to adjust VE results for quantity of smoking among former and current smokers.

1.16 V-3. Effect Modification – Example 2

	In this study, wwo interaction terms, (current smoking X middle VE and current smoking X high VE) were also included. By definition, synergism means that the combined effects of two factors on an outcome is greater than the *sum* of the two individual effects, not necessarily the product (36). Rothman and others have argued that regardless of model, interaction between two risk factors should be assessed on an additive scale as the impact in terms of number of excess cases (attributable risk) depends on the risk difference rather than the relative risk. (37, 38). Thus, this article assesses the presence of interaction on the additive scale and employs the relative excess risk due to interaction (RERI) calculation as a summary statistic. RERI represents risk that is in excess of what would be expected if the combination of two risk factors resulted in interaction of a purely additive nature. In a proportional hazards model, testing the hypothesis that the β coefficient of the interaction term is zero is testing that the HR for those with both

	factors is equal to the product of the HR for the first factor *times* the HR for the second factor, which is a greater quantity than the sum. In terms of the model coefficients, RERI is calculated as e(β1+ β2+β3) -e(β1) -e(β2) + 1 where e denotes the exponent, and β1 , β2, and β3 are the coefficients from the model for specified levels of VE and smoking, and their interaction respectively. We tested the hypothesis RERI = 0 by a z-test (normal distribution). We then expressed RERI as a percent of total excess risk when both factors are present (i.e. RERI% ={RERI/[HR(AB) - 1]} X 100. RERI% is the proportion of disease burden caused by two factors that can be attributed to their interaction. We see that the RERI of 81% represents the excess burden of disease associated with current smoking and VE is attributed to their interaction. For people with moderate VE, this excess risk is larger at 93%.

1.17 What we will cover this unit:

	This concludes our review of a few examples presenting logistic regression data. I suggest you take some time to review these articles we've highlighted. As an epidemiologist, you will be expected to run similar anlayses of data of your own, and present them in a similar manner.