2 Important Special Cases of the Logistic Model

Contents

Introduction 42
Abbreviated Outline 42
Objectives 43
Presentation 45
Detailed Outline 65
Practice Exercises 67
Test 69
Answers to Practice Exercises 71
2. Important Special Cases of the Logistic Model

**Introduction**

In this chapter, several important special cases of the logistic model involving a single (0, 1) exposure variable are considered with their corresponding odds ratio expressions. In particular, focus is on defining the independent variables that go into the model and on computing the odds ratio for each special case. Models that account for the potential confounding effects and potential interaction effects of covariates are emphasized.

**Abbreviated Outline**

The outline below gives the user a preview of the material to be covered by the presentation. A detailed outline for review purposes follows the presentation.

I. **Overview** (page 45)

II. **Special case – Simple analysis** (pages 46–49)

III. **Assessing multiplicative interaction** (pages 49–55)

IV. **The E, V, W model** – A general model containing a (0, 1) exposure and potential confounders and effect modifiers (pages 55–64)
Objectives

Upon completion of this chapter, the learner should be able to:

1. State or recognize the logistic model for a simple analysis.
2. Given a model for simple analysis:
   a. state an expression for the odds ratio describing the exposure–disease relationship
   b. state or recognize the null hypothesis of no exposure–disease relationship in terms of parameter(s) of the model
   c. compute or recognize an expression for the risk for exposed or unexposed persons separately
   d. compute or recognize an expression for the odds of getting the disease for exposed or unexposed persons separately
3. Given two (0, 1) independent variables:
   a. state or recognize a logistic model that allows for the assessment of interaction on a multiplicative scale
   b. state or recognize the expression for no interaction on a multiplicative scale in terms of odds ratios for different combinations of the levels of two (0, 1) independent variables
   c. state or recognize the null hypothesis for no interaction on a multiplicative scale in terms of one or more parameters in an appropriate logistic model
4. Given a study situation involving a (0, 1) exposure variable and several control variables:
   a. state or recognize a logistic model that allows for the assessment of the exposure-disease relationship, controlling for the potential confounding and potential interaction effects of functions of the control variables
   b. compute or recognize the expression for the odds ratio for the effect of exposure on disease status adjusting for the potential confounding and interaction effects of the control variables in the model
   c. state or recognize an expression for the null hypothesis of no interaction effect involving one or more of the effect modifiers in the model
   d. assuming no interaction, state or recognize an expression for the odds ratio for the effect of exposure on disease status adjusted for confounders
2. Important Special Cases of the Logistic Model

  e. assuming no interaction, state or recognize the null hypothesis for testing the significance of this odds ratio in terms of a parameter in the model

  5. Given a logistic model involving interaction terms, state or recognize that the expression for the odds ratio will give different values for the odds ratio depending on the values specified for the effect modifiers in the model.
This presentation describes important special cases of the general logistic model when there is a single (0, 1) exposure variable. Special case models include simple analysis of a fourfold table, assessment of multiplicative interaction between two dichotomous variables, and controlling for several confounders and interaction terms. In each case, we consider the definitions of variables in the model and the formula for the odds ratio describing the exposure-disease relationship.

Recall that the general logistic model for \( k \) independent variables may be written as \( P(X) = \frac{1}{1 + e^{-(\alpha + \sum \beta_i X_i)}} \) where \( P(X) \) denotes the probability of developing a disease of interest given values of a collection of independent variables \( X_1, X_2, \ldots, X_k \), that are collectively denoted by the \textit{bold} \( X \). The terms \( \alpha \) and \( \beta_i \) in the model represent unknown parameters that we need to estimate from data obtained for a group of subjects on the \( X \)s and on \( D \), a dichotomous disease outcome variable.

An alternative way of writing the logistic model is called the logit form of the model. The expression for the logit form is given here.

The general odds ratio formula for the logistic model is given by either of two formulae. The first formula is of the form \( e^{\alpha + \sum \beta_i X_i} \) equals 1 over 1 plus the quantity \( \alpha \) plus the sum of \( \beta_i X_i \), where \( P(X) \) denotes the probability of developing a disease of interest given values of a collection of independent variables \( X_1, X_2, \ldots, X_k \), that are collectively denoted by the \textit{bold} \( X \). The terms \( \alpha \) and \( \beta_i \) in the model represent unknown parameters that we need to estimate from data obtained for a group of subjects on the \( X \)s and on \( D \), a dichotomous disease outcome variable.

We now consider a number of important special cases of the logistic model and their corresponding odds ratio formulae.
II. Special Case – Simple Analysis

We begin with the simple situation involving one dichotomous independent variable, which we will refer to as an exposure variable and will denote it as $X_1 = E$. Because the disease variable, $D$, considered by a logistic model is dichotomous, we can use a two-way table with four cells to characterize this analysis situation, which is often referred to as a simple analysis.

For convenience, we define the exposure variable as a $(0, 1)$ variable and place its values in the two columns of the table. We also define the disease variable as a $(0, 1)$ variable and place its values in the rows of the table. The cell frequencies within the fourfold table are denoted as $a$, $b$, $c$, and $d$, as is typically presented for such a table.

A logistic model for this simple analysis situation can be defined by the expression $P(X) = \frac{1}{1 + e^{-(x + \beta_1 E)}}$, where $E = (0, 1)$ variable.

Note: Other coding schemes $(1, -1), (1, 2), (2, 1)$

The logit form of the logistic model we have just defined is of the form $\logit P(X) = x + \beta_1 E$.

The term $P(X)$ for the simple analysis model denotes the probability that the disease variable $D$ takes on the value 1, given whatever the value is for the exposure variable $E$. In epidemiologic terms, this probability denotes the risk for developing the disease, given exposure status. When the value of the exposure variable equals 1, we call this risk $R_1$, which is the conditional probability that $D$ equals 1 given that $E$ equals 1. When $E$ equals 0, we denote the risk by $R_0$, which is the conditional probability that $D$ equals 1 given that $E$ equals 0.
We would like to use the above model for simple analysis to obtain an expression for the odds ratio that compares exposed persons with unexposed persons. Using the terms $R_1$ and $R_0$, we can write this odds ratio as $R_1$ divided by 1 minus $R_1$ over $R_0$ divided by 1 minus $R_0$.

To compute the odds ratio in terms of the parameters of the logistic model, we substitute the logistic model expression into the odds ratio formula.

For $E$ equal to 1, we can write $R_1$ by substituting the value $E$ equals 1 into the model formula for $P(X)$. We then obtain 1 over 1 plus $e$ to minus the quantity $z$ plus $\beta_1$ times 1, or simply 1 over 1 plus $e$ to minus $z$ plus $\beta_1$.

For $E$ equal to zero, we write $R_0$ by substituting $E$ equal to 0 into the model formula, and we obtain 1 over 1 plus $e$ to minus $z$.

To obtain ROR then, we replace $R_1$ with 1 over 1 plus $e$ to minus $z$ plus $\beta_1$, and we replace $R_0$ with 1 over 1 plus $e$ to minus $z$. The ROR formula then simplifies algebraically to $e$ to the $\beta_1$, where $\beta_1$ is the coefficient of the exposure variable.

We could have obtained this expression for the odds ratio using the general formula for the ROR that we gave during our review. We will use the general formula now. Also, for other special cases of the logistic model, we will use the general formula rather than derive an odds ratio expression separately for each case.
2. Important Special Cases of the Logistic Model

The general formula computes ROR as $e$ to the sum of each $\beta_i$ times the difference between $X_{1i}$ and $X_{0i}$, where $X_{1i}$ denotes the value of the $i$th $X$ variable for group 1 persons and $X_{0i}$ denotes the value of the $i$th $X$ variable for group 0 persons. In a simple analysis, we have only one $X$ and one $\beta_i$; in other words, $k$, the number of variables in the model, equals 1.

For a simple analysis model, group 1 corresponds to exposed persons, for whom the variable $X_1$, in this case $E$, equals 1. Group 0 corresponds to unexposed persons, for whom the variable $X_1$ or $E$ equals 0. Stated another way, for group 1, the collection of $X$s denoted by the bold $X$ can be written as $X_1$ and equals the collection of one value $X_{11}$, which equals 1. For group 0, the collection of $X$s denoted by the bold $X$ is written as $X_0$ and equals the collection of one value $X_{01}$, which equals 0.

Substituting the particular values of the one $X$ variable into the general odds ratio formula then gives $e$ to the $\beta_1$ times the quantity $X_{11}$ minus $X_{01}$, which becomes $e$ to the $\beta_1$ times 1 minus 0, which reduces to $e$ to the $\beta_1$.

In summary, for the simple analysis model involving a (0, 1) exposure variable, the logistic model $P(X)$ equals 1 over 1 plus $e$ to minus the quantity $z$ plus $\beta_1$ times $E$, and the odds ratio that describes the effect of the exposure variable is given by $e$ to the $\beta_1$, where $\beta_1$ is the coefficient of the exposure variable.

We can estimate this odds ratio by fitting the simple analysis model to a set of data. The estimate of the parameter $\beta_1$ is typically denoted as $\hat{\beta}_1$. The odds ratio estimate then becomes $e$ to the $\hat{\beta}_1$. 

\[
\text{SIMPLE ANALYSIS SUMMARY}
\]

\[
P(X) = \frac{1}{1 + e^{-(z + \beta_1 E)}}
\]

\[
\text{ROR} = e^{\beta_1}
\]
### III. Assessing Multiplicative Interaction

Consider, for example, two (0, 1) $X$ variables, $X_1$ and $X_2$, which for convenience we rename as $A$ and $B$, respectively. We first describe what we mean conceptually by interaction between these two variables. This involves an equation involving risk odds ratios corresponding to different combinations of $A$ and $B$. The odds ratios are defined in terms of risks, which we now describe.

Let $R_{AB}$ denote the risk for developing the disease, given specified values for $A$ and $B$; in other words, $R_{AB}$ equals the conditional probability that $D$ equals 1, given $A$ and $B$.

Because $A$ and $B$ are dichotomous, there are four possible values for $R_{AB}$, which are shown in the cells of a two-way table. When $A$ equals 1 and $B$ equals 1, the risk $R_{AB}$ becomes $R_{11}$. Similarly, when $A$ equals 1 and $B$ equals 0, the risk becomes $R_{10}$. When $A$ equals 0 and $B$ equals 1, the risk is $R_{01}$, and finally, when $A$ equals 0 and $B$ equals 0, the risk is $R_{00}$.

The reader should not be surprised to find out that an alternative formula for the estimated odds ratio for the simple analysis model is the familiar $a$ times $d$ over $b$ times $c$, where $a$, $b$, $c$, and $d$ are the cell frequencies in the fourfold table for simple analysis. That is, $e$ to the $\beta_1$ obtained from fitting a logistic model for simple analysis can alternatively be computed as $ad$ divided by $bc$ from the cell frequencies of the fourfold table.

Thus, in the simple analysis case, we need not go to the trouble of fitting a logistic model to get an odds ratio estimate as the typical formula can be computed without a computer program. We have presented the logistic model version of simple analysis to show that the logistic model incorporates simple analysis as a special case. More complicated special cases, involving more than one independent variable, require a computer program to compute the odds ratio.

We will now consider how the logistic model allows the assessment of interaction between two independent variables.
Note: above table not for simple analysis.

<table>
<thead>
<tr>
<th>A</th>
<th>B = 1</th>
<th>B = 0</th>
</tr>
</thead>
<tbody>
<tr>
<td>A = 1</td>
<td>( R_{11} )</td>
<td>( R_{10} )</td>
</tr>
<tr>
<td>A = 0</td>
<td>( R_{01} )</td>
<td>( R_{00} )</td>
</tr>
</tbody>
</table>

Note that the two-way table presented here does not describe a simple analysis because the row and column headings of the table denote two independent variables rather than one independent variable and one disease variable. Moreover, the information provided within the table is a collection of four risks corresponding to different combinations of both independent variables, rather than four cell frequencies corresponding to different exposure-disease combinations.

Within this framework, odds ratios can be defined to compare the odds for any one cell in the two-way table of risks with the odds for any other cell. In particular, three odds ratios of typical interest compare each of three of the cells to a referent cell. The referent cell is usually selected to be the combination \( A = 0 \) and \( B = 0 \). The three odds ratios are then defined as \( \text{OR}_{11} \), \( \text{OR}_{10} \), and \( \text{OR}_{01} \), where \( \text{OR}_{11} \) equals the odds for cell 11 divided by the odds for cell 00, \( \text{OR}_{10} \) equals the odds for cell 10 divided by the odds for cell 00, and \( \text{OR}_{01} \) equals the odds for cell 01 divided by the odds for cell 00.

As the odds for any cell \( A, B \) is defined in terms of risks as \( R_{AB} \) divided by 1 minus \( R_{AB} \), we can obtain the following expressions for the three odds ratios: \( \text{OR}_{11} \) equals the product of \( R_{11} \) times 1 minus \( R_{00} \) divided by the product of \( R_{00} \) and 1 minus \( R_{11} \). The corresponding expressions for \( \text{OR}_{10} \) and \( \text{OR}_{01} \) are similar, where the subscript 11 in the numerator and denominator of the 11 formula is replaced by 10 and 01, respectively.

In general, without specifying the value of \( A \) and \( B \), we can write the odds ratio formulae as \( \text{OR}_{AB} \) equals the product of \( R_{AB} \) and 1 minus \( R_{00} \) divided by the product of \( R_{00} \) and 1 minus \( R_{AB} \), where \( A \) takes on the values 0 and 1 and \( B \) takes on the values 0 and 1.
DEFINITION

\[ OR_{11} = OR_{10} \times OR_{01} \]

no interaction on a multiplicative scale

Now that we have defined appropriate odds ratios for the two independent variables situation, we are ready to provide an equation for assessing interaction. The equation is stated as \( OR_{11} \) equals the product of \( OR_{10} \) and \( OR_{01} \). If this expression is satisfied for a given study situation, we say that there is “no interaction on a multiplicative scale.” In contrast, if this expression is not satisfied, we say that there is evidence of interaction on a multiplicative scale.

Note that the right-hand side of the “no interaction” expression requires multiplication of two odds ratios, one corresponding to the combination 10 and the other to the combination 01. Thus, the scale used for assessment of interaction is called multiplicative.

No interaction:

\[
\begin{align*}
\text{(effect of } A \text{ and } B \text{ acting together)} & = \text{(combined effect of } A \text{ and } B \text{ acting separately)} \\
OR_{11} & = OR_{10} \times OR_{01} \text{ multiplicative scale}
\end{align*}
\]

When the no interaction equation is satisfied, we can interpret the effect of both variables \( A \) and \( B \) acting together as being the same as the combined effect of each variable acting separately.

The effect of both variables acting together is given by the odds ratio \( OR_{11} \) obtained when \( A \) and \( B \) are both present, that is, when \( A = 1 \) and \( B = 1 \).

The effect of \( A \) acting separately is given by the odds ratio for \( A = 1 \) and \( B = 0 \), and the effect of \( B \) acting separately is given by the odds ratio for \( A = 0 \) and \( B = 1 \). The combined separate effects of \( A \) and \( B \) are then given by the product \( OR_{10} \) times \( OR_{01} \).

Thus, when there is no interaction on a multiplicative scale, \( OR_{11} \) equals the product of \( OR_{10} \) and \( OR_{01} \).
As an example of no interaction on a multiplicative scale, suppose the risks $R_{AB}$ in the four-fold table are given by $R_{11}$ equal to 0.0350, $R_{10}$ equal to 0.0175, $R_{01}$ equal to 0.0050, and $R_{00}$ equal to 0.0025. Then the corresponding three odds ratios are obtained as follows: $OR_{11}$ equals 0.0350 times 1 minus 0.0025 divided by the product of 0.0025 and 1 minus 0.0350, which becomes 14.4; $OR_{10}$ equals 0.0175 times 1 minus 0.0025 divided by the product of 0.0025 and 1 minus 0.0175, which becomes 7.2; and $OR_{01}$ equals 0.0050 times 1 minus 0.0025 divided by the product of 0.0025 and 1 minus 0.0050, which becomes 2.0.

To see if the no interaction equation is satisfied, we check whether $OR_{11}$ equals the product of $OR_{10}$ and $OR_{01}$. Here we find that $OR_{11}$ equals 14.4 and the product of $OR_{10}$ and $OR_{01}$ is 7.2 times 2, which is also 14.4. Thus, the no interaction equation is satisfied.

In contrast, using a different example, if the risk for the 11 cell is 0.0700, whereas the other three risks remained at 0.0175, 0.0050, and 0.0025, then the corresponding three odds ratios become $OR_{11}$ equals 30.0, $OR_{10}$ equals 7.2, and $OR_{01}$ equals 2.0. In this case, the no interaction equation is not satisfied because the left-hand side equals 30 and the product of the two odds ratios on the right-hand side equals 14. Here, then, we would conclude that there is interaction because the effect of both variables acting together is more than twice the combined effect of the variables acting separately.
Logistic model variables:

\[ X_1 = A_{(0,1)} \quad \text{main effects} \]
\[ X_2 = B_{(0,1)} \]
\[ X_3 = A \times B \quad \text{interaction effect variable} \]

\[
\logit P(X) = \alpha + \beta_1 A + \beta_2 B + \beta_3 A \times B,
\]

where

\[
P(X) = \text{risk given } A \text{ and } B \]
\[
= R_{AB}
\]

\[
\beta_3 = \ln \left[ \frac{OR_{11}}{OR_{10} \times OR_{01}} \right]
\]

A more complete discussion of interaction, including the distinction between multiplicative interaction and additive interaction, is given in Chap. 19 of *Epidemiologic Research* by Kleinbaum, Kupper, and Morgenstern (1982).

We now define a logistic model that allows the assessment of multiplicative interaction involving two \((0, 1)\) indicator variables \(A\) and \(B\). This model contains three independent variables, namely, \(X_1\) equal to \(A\), \(X_2\) equal to \(B\), and \(X_3\) equal to the product term \(A\) times \(B\). The variables \(A\) and \(B\) are called main effect variables and the product term is called an interaction effect variable.

The logit form of the model is given by the expression \(\logit P(X)\) equals \(\alpha\) plus \(\beta_1\) times \(A\) plus \(\beta_2\) times \(B\) plus \(\beta_3\) times \(A\) times \(B\). \(P(X)\) denotes the risk for developing the disease given values of \(A\) and \(B\), so that we can alternatively write \(P(X)\) as \(R_{AB}\).

For this model, it can be shown mathematically that the coefficient \(\beta_3\) of the product term can be written in terms of the three odds ratios we have previously defined. The formula is \(\beta_3\) equals the natural log of the quantity \(OR_{11}\) divided by the product of \(OR_{10}\) and \(OR_{01}\). We can make use of this formula to test the null hypothesis of no interaction on a multiplicative scale.
H₀ no interaction on a multiplicative scale
\[ H₀ : \frac{\text{OR}_{11}}{\text{OR}_{10} \times \text{OR}_{01}} = 1 \]
\[ H₀ : \ln \left( \frac{\text{OR}_{11}}{\text{OR}_{10} \times \text{OR}_{01}} \right) = \ln 1 \]
\[ H₀ : \beta_3 = 0 \]

logit \( \text{P(X)} = \alpha + \beta_1 A + \beta_2 B + \beta_3 AB \)
H₀: no interaction \( \Leftrightarrow \beta_3 = 0 \)

Test result Model
not significant \( \Rightarrow \alpha + \beta_1 A + \beta_2 B \)
significant \( \Rightarrow \alpha + \beta_1 A + \beta_2 B + \beta_3 AB \)

MAIN POINT:
Interaction test \( \Rightarrow \) test for product terms

A description of methods for testing hypotheses for logistic regression models is beyond the scope of this presentation (see Chap. 5). The main point here is that we can test for interaction in a logistic model by testing for significance of product terms that reflect interaction effects in the model.

As an example of a test for interaction, we consider a study that looks at the combined relationship of asbestos exposure and smoking to the development of bladder cancer. Suppose we have collected case-control data on several persons with the same occupation. We let ASB denote a \((0, 1)\) variable indicating asbestos exposure status, SMK denote a \((0, 1)\) variable indicating smoking status, and \(D\) denote a \((0, 1)\) variable for bladder cancer status.

**EXAMPLE**

<table>
<thead>
<tr>
<th>Case-control study</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ASB = ((0, 1))</td>
<td>variable for asbestos exposure</td>
</tr>
<tr>
<td>SMK = ((0, 1))</td>
<td>variable for smoking status</td>
</tr>
<tr>
<td>(D) = ((0, 1))</td>
<td>variable for bladder cancer status</td>
</tr>
</tbody>
</table>
IV. The E, V, W Model – A General Model Containing a (0, 1) Exposure and Potential Confounders and Effect Modifiers

The variables:

\[ E = (0, 1) \text{ exposure} \]

\[ C_1, C_2, \ldots, C_p \text{ continuous or categorical} \]

To assess the extent to which there is a multiplicative interaction between asbestos exposure and smoking, we consider a logistic model with ASB and SMK as main effect variables and the product term ASB times SMK as an interaction effect variable. The model is given by the expression

\[
\text{logit } P(X) = \alpha + \beta_1 \text{ASB} + \beta_2 \text{SMK} + \beta_3 \text{ASB} \times \text{SMK}
\]

The null hypothesis is

\[ H_0: \beta_3 = 0 \]

If this test is not significant, then we would conclude that the effect of asbestos and smoking acting together is equal, on a multiplicative scale, to the combined effect of asbestos and smoking acting separately. If this test is significant and \( \beta_3 > 0 \), we would conclude that the joint effect of asbestos and smoking is greater than a multiplicative combination of separate effects. Or, if the test is significant and \( \beta_3 < 0 \), we would conclude that the joint effect of asbestos and smoking is less than a multiplicative combination of separate effects.

We are now ready to discuss a logistic model that considers the effects of several independent variables and, in particular, allows for the control of confounding and the assessment of interaction. We call this model the E, V, W model. We consider a single dichotomous (0, 1) exposure variable, denoted by E, and p extraneous variables \( C_1, C_2, \ldots, C_p \), and so on, up through \( C_p \). The variables \( C_1 \) through \( C_p \) may be either continuous or categorical.

As an example of this special case, suppose the disease variable is coronary heart disease status (CHD), the exposure variable \( E \) is catecholamine level (CAT), where 1 equals high and 0 equals low, and the control variables are AGE, cholesterol level (CHL), smoking status (SMK), electrocardiogram abnormality status (ECG), and hypertension status (HPT).
The general $E, V, W$ Model

single exposure, controlling for $C_1, C_2, \ldots, C_p$

We will assume here that both AGE and CHL are treated as continuous variables, that SMK is a $(0, 1)$ variable, where 1 equals ever smoked and 0 equals never smoked, that ECG is a $(0, 1)$ variable, where 1 equals abnormality present and 0 equals abnormality absent, and that HPT is a $(0, 1)$ variable, where 1 equals high blood pressure and 0 equals normal blood pressure. There are, thus, five $C$ variables in addition to the exposure variable CAT.

We now consider a model with eight independent variables. In addition to the exposure variable CAT, the model contains the five $C$ variables as potential confounders plus two product terms involving two of the $C$s, namely, CHL and HPT, which are each multiplied by the exposure variable CAT.

The model is written as logit $P(X) = \alpha + \beta \text{CAT}$ plus the sum of five main effect terms $\gamma_1 \text{AGE} + \gamma_2 \text{CHL} + \gamma_3 \text{SMK} + \gamma_4 \text{ECG} + \gamma_5 \text{HPT}$ and the sum of $\delta_1 \text{CAT} \times \text{CHL}$ and $\delta_2 \text{CAT} \times \text{HPT}$.

Parameters:

- $\alpha$: constant
- $\beta$: exposure variable
- $\gamma$s: potential confounders
- $\delta$s: potential interaction variables

Note that the parameters in this model are denoted as $\beta, \gamma$s, and $\delta$s instead of $a$ and $b$s, whereas previously we denoted all parameters other than the constant $\alpha$ as $\beta$s. We use $\beta, \gamma$s, and $\delta$s here to distinguish different types of variables in the model. The parameter $\beta$ indicates the coefficient of the exposure variable, the $\gamma$s indicate the coefficients of the potential confounders in the model, and the $\delta$s indicate the coefficients of the potential interaction variables in the model. This notation for the parameters will be used throughout the remainder of this presentation.

Analogous to the above example, we now describe the general form of a logistic model, called the $E, V, W$ model, that considers the effect of a single exposure controlling for the potential confounding and interaction effects of control variables $C_1, C_2, \ldots, C_p$. 
$E, V, W$ Model

\[ k = p_1 + p_2 + 1 = \text{no. of variables in model} \]
\[ p_1 = \text{no. of potential confounders} \]
\[ p_2 = \text{no. of potential interactions} \]
\[ 1 = \text{exposure variable} \]

The general $E, V, W$ model contains $p_1$ plus $p_2$ plus 1 variables, where $p_1$ is the number of potential confounders in the model, $p_2$ is the number of potential interaction terms in the model, and 1 denotes the exposure variable.

In the CHD study example above, there are $p_1$ equals to five potential confounders, namely, the five control variables, and there are $p_2$ equal to two interaction variables, the first of which is $\text{CAT} \times \text{CHL}$ and the second is $\text{CAT} \times \text{HPT}$. The total number of variables in the example is, therefore, $p_1$ plus $p_2$ plus 1 equals 5 plus 2 plus 1, which equals 8. This corresponds to the model presented earlier, which contained eight variables.

In addition to the exposure variable $E$, the general model contains $p_1$ variables denoted as $V_1$ through $V_{p_1}$. The set of $V$s are functions of the Cs that are thought to account for confounding in the data. We call the set of these $V$s potential confounders.

For instance, we may have $V_1$ equal to $C_1$, $V_2$ equal to $(C_2)^2$, and $V_3$ equal to $C_1 \times C_3$.

The CHD example above has five $V$s that are the same as the Cs.

Following the $V$s, we define $p_2$ variables that are product terms of the form $E$ times $W_1$, $E$ times $W_2$, and so on up through $E$ times $W_{p_2}$, where $W_1$, $W_2$, through $W_{p_2}$, denote a set of functions of the Cs that are potential effect modifiers with $E$.

For instance, we may have $W_1$ equal to $C_1$ and $W_2$ equal to $C_1 \times C_3$.

The CHD example above has two $W$s, namely, CHL and HPT, that go into the model as product terms of the form $\text{CAT} \times \text{CHL}$ and $\text{CAT} \times \text{HPT}$.

<table>
<thead>
<tr>
<th>CHD EXAMPLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_1 = 5$: AGE, CHL, SMK, ECG, HPT</td>
</tr>
<tr>
<td>$p_2 = 2$: CAT $\times$ CHL, CAT $\times$ HPT</td>
</tr>
<tr>
<td>$p_1 + p_2 + 1 = 5 + 2 + 1 = 8$</td>
</tr>
</tbody>
</table>

- $V_1, \ldots, V_{p_1}$ are potential confounders
- $V$s are functions of Cs

\[ e.g., V_1 = C_1, V_2 = (C_2)^2, V_3 = C_1 \times C_3 \]

<table>
<thead>
<tr>
<th>CHD EXAMPLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V_1 = \text{AGE}$, $V_2 = \text{CHL}$, $V_3 = \text{SMK}$, $V_4 = \text{ECG}$, $V_5 = \text{HPT}$</td>
</tr>
</tbody>
</table>

- $W_1, \ldots, W_{p_2}$ are potential effect modifiers
- $W$s are functions of Cs

\[ e.g., W_1 = C_1, W_2 = C_1 \times C_3 \]

<table>
<thead>
<tr>
<th>CHD EXAMPLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>$W_1 = \text{CHL}$, $W_2 = \text{HPT}$</td>
</tr>
</tbody>
</table>
REFERENCES FOR CHOICE OF Vs AND Ws FROM Cs

- Chap. 6: Modeling Strategy Guidelines
- Epidemiologic Research, Chap. 21

Assume: Vs and Ws are Cs or subset of Cs

NOTE
Ws ARE SUBSET OF Vs

EXAMPLE
\[ C_1 = \text{AGE}, C_2 = \text{RACE}, C_3 = \text{SEX} \]
\[ V_1 = \text{AGE}, V_2 = \text{RACE}, V_3 = \text{SEX} \]
\[ W_1 = \text{AGE}, W_2 = \text{SEX} \]
\[ p_1 = 3, p_2 = 2, k = p_1 + p_2 + 1 = 6 \]

A logistic model incorporating this special case containing the E, V, and W variables defined above can be written in logit form as shown here.

\[
\text{logit P}(X) = \alpha + \beta E + \gamma_1 V_1 + \gamma_2 V_2 + \cdots + \gamma_{p_1} V_{p_1} + \delta_1 E W_1 + \delta_2 E W_2 + \cdots + \delta_{p_2} E W_{p_2},
\]

where
\[ \beta = \text{coefficient of } E \]
\[ \gamma_i = \text{coefficient of } V_i \]
\[ \delta_j = \text{coefficient of } W_j \]

We can factor out the E from each of the interaction terms, so that the model may be more simply written as shown here. This is the form of the model that we will use henceforth in this presentation.

EXAMPLE
\[ V_1 = \text{AGE}, V_2 = \text{RACE} \]
\[ W_1 = \text{AGE}, W_2 = \text{SEX} \]
Adjusted odds ratio for $E = 1$ vs. $E = 0$ given $C_1, C_2, \ldots, C_p$ fixed

$\text{ROR} = \exp\left( \beta + \sum_{j=1}^{p_2} \delta_j W_j \right)$

We now provide for this model an expression for an adjusted odds ratio that describes the effect of the exposure variable on disease status adjusted for the potential confounding and interaction effects of the control variables $C_1$ through $C_p$. That is, we give a formula for the risk odds ratio comparing the odds of disease development for exposed vs. unexposed persons, with both groups having the same values for the extraneous factors $C_1$ through $C_p$. This formula is derived as a special case of the odds ratio formula for a general logistic model given earlier in our review.

For our special case, the odds ratio formula takes the form $\text{ROR} = e^{\beta + \sum_{j=1}^{p_2} \delta_j W_j}$

Note that $\beta$ is the coefficient of the exposure variable $E$, that the $\delta_j$ are the coefficients of the interaction terms of the form $E$ times $W_j$, and that the coefficients $\gamma_i$ of the main effect variables $V_i$ do not appear in the odds ratio formula.

Note also that this formula assumes that the dichotomous variable $E$ is coded as a (0, 1) variable with $E$ equal to 1 for exposed persons and $E$ equal to 0 for unexposed persons. If the coding scheme is different, for example, (1, -1) or (2, 1), or if $E$ is an ordinal or interval variable, then the odds ratio formula needs to be modified. The effect of different coding schemes on the odds ratio formula will be described in Chap. 3.

This odds ratio formula tells us that if our model contains interaction terms, then the odds ratio will involve coefficients of these interaction terms and that, moreover, the value of the odds ratio will be different depending on the values of the $W$ variables involved in the interaction terms as products with $E$. This property of the OR formula should make sense in that the concept of interaction implies that the effect of one variable, in this case $E$, is different at different levels of another variable, such as any of the $W$s.
- Vs not in OR formula but Vs in model, so OR formula controls confounding:

\[
\text{logit } P(X) = \alpha + \beta E + \sum (\gamma_i)V_i + E \sum (\delta_j)W_j
\]

No interaction:

all \( \delta_j = 0 \) \( \Rightarrow \) \( \text{ROR} = \exp(\beta) \) \( \uparrow \)

constant

\[
\text{logit } P(X) = \alpha + \beta E + \sum \gamma_iV_i \uparrow \text{confounding effects adjusted}
\]

Although the coefficients of the \( V \) terms do not appear in the odds ratio formula, these terms are still part of the fitted model. Thus, the odds ratio formula not only reflects the interaction effects in the model but also controls for the confounding variables in the model.

In contrast, if the model contains no interaction terms, then, equivalently, all the \( \delta_j \) coefficients are 0; the odds ratio formula thus reduces to \( \text{ROR} = e^\beta \), where \( \beta \) is the coefficient of the exposure variable \( E \). Here, the odds ratio is a fixed constant, so that its value does not change with different values of the independent variables. The model in this case reduces to logit \( P(X) \) equals \( \alpha \) plus \( \beta \) times \( E \) plus the sum of the main effect terms involving the \( V \)s and contains no product terms. For this model, we can say that \( e^\beta \) represents an odds ratio that adjusts for the potential confounding effects of the control variables \( C_1 \) through \( C_p \) defined in terms of the \( V \)s.

As an example of the use of the odds ratio formula for the \( E, V, W \) model, we return to the CHD study example we described earlier. The CHD study model contained eight independent variables. The model is restated here as logit \( P(X) \) equals \( \alpha \) plus \( \beta \) times \( CAT \) plus the sum of five main effect terms plus the sum of two interaction terms.

The five main effect terms in this model account for the potential confounding effects of the variables \( AGE \) through \( HPT \). The two product terms account for the potential interaction effects of \( CHL \) and \( HPT \) with \( CAT \).

For this example, the odds ratio formula reduces to the expression \( \text{ROR} = e^\beta \) plus the sum \( \delta_1 \) times \( CHL \) plus \( \delta_2 \) times \( HPT \).
In using this formula, note that to obtain a numerical value for this odds ratio, not only do we need estimates of the coefficients $\beta$ and the two $\delta$s, but we also need to specify values for the variables CHL and HPT. In other words, once we have fitted the model to obtain estimates of the coefficients, we will get different values for the odds ratio depending on the values that we specify for the interaction variables in our model. Note, also, that although the variables AGE, SMK, and ECG are not contained in the odds ratio expression for this model, the confounding effects of these three variables plus CHL and HPT are being adjusted because the model being fit contains all five control variables as main effect $V$ terms.

To provide numerical values for the above odds ratio, we will consider a data set of 609 white males from Evans County, Georgia, who were followed for 9 years to determine CHD status. The above model involving CAT, the five $V$ variables, and the two $W$ variables was fit to this data, and the fitted model is given by the list of coefficients corresponding to the variables listed here.

Based on the above fitted model, the estimated odds ratio for the CAT, CHD association adjusted for the five control variables is given by the expression shown here. Note that this expression involves only the coefficients of the exposure variable CAT and the interaction variables CAT times CHL and CAT times HPT, the latter two coefficients being denoted by $\delta$s in the model.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>$\hat{a}$ = -4.0497</td>
</tr>
<tr>
<td>CAT</td>
<td>$\hat{\beta}$ = -12.6894</td>
</tr>
<tr>
<td>AGE</td>
<td>$\hat{\gamma}_1$ = 0.0350</td>
</tr>
<tr>
<td>CHL</td>
<td>$\hat{\gamma}_2$ = -0.0055</td>
</tr>
<tr>
<td>SMK</td>
<td>$\hat{\gamma}_3$ = 0.7732</td>
</tr>
<tr>
<td>ECG</td>
<td>$\hat{\gamma}_4$ = 0.3671</td>
</tr>
<tr>
<td>HPT</td>
<td>$\hat{\gamma}_5$ = 1.0466</td>
</tr>
<tr>
<td>CAT $\times$ CHL</td>
<td>$\hat{\delta}_1$ = 0.0692</td>
</tr>
<tr>
<td>CAT $\times$ HPT</td>
<td>$\hat{\delta}_2$ = -2.3318</td>
</tr>
</tbody>
</table>

\[ \text{ROR} = \exp\left(-12.6894 + 0.0692\text{CHL} - 2.3318\text{HPT}\right) \]

In the model, $\hat{a}$ varies with values of CHL and HPT. AGE, SMK, and ECG are adjusted for confounding.
Choice of $W$ values depends on investigator

**EXAMPLE**

<table>
<thead>
<tr>
<th>TABLE OF POINT ESTIMATES $ROR$</th>
<th>HPT = 0</th>
<th>HPT = 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHL = 180</td>
<td>0.79</td>
<td>0.08</td>
</tr>
<tr>
<td>CHL = 200</td>
<td>3.16</td>
<td>0.31</td>
</tr>
<tr>
<td>CHL = 220</td>
<td>12.61</td>
<td>1.22</td>
</tr>
<tr>
<td>CHL = 240</td>
<td>50.33</td>
<td>4.89</td>
</tr>
</tbody>
</table>

**EXAMPLE (continued)**

This expression for the odds ratio tells us that we obtain a different value for the estimated odds ratio depending on the values specified for CHL and HPT. As previously mentioned, this should make sense conceptually because CHL and HPT are the only two effect modifiers in the model, and the value of the odds ratio changes as the values of the effect modifiers change.

To get a numerical value for the odds ratio, we consider, for example, the specific values CHL equal to 220 and HPT equal to 1. Plugging these into the odds ratio formula, we obtain $e$ to the 0.2028, which equals 1.22.

As a second example, we consider CHL equal to 200 and HPT equal to 0. Here, the odds ratio becomes $e$ to 1.1506, which equals 3.16.

Thus, we see that depending on the values of the effect modifiers we will get different values for the estimated odds ratios. Note that each estimated odds ratio obtained adjusts for the confounding effects of all five control variables because these five variables are contained in the fitted model as $V$ variables.

In general, when faced with an odds ratio expression involving effect modifiers ($W$), the choice of values for the $W$ variables depends primarily on the interest of the investigator. Typically, the investigator will choose a range of values for each interaction variable in the odds ratio formula; this choice will lead to a table of estimated odds ratios, such as the one presented here, for a range of CHL values and the two values of HPT. From such a table, together with a table of confidence intervals, the investigator can interpret the exposure–disease relationship.

As a second example, we consider a model containing no interaction terms from the same Evans County data set of 609 white males. The variables in the model are the exposure variable CAT, and five $V$ variables, namely, AGE, CHL, SMK, ECG, and HPT. This model is written in logit form as shown here.
Because this model contains no interaction terms, the odds ratio expression for the CAT, CHD association is given by $e^{\hat{\beta}}$, where $\hat{\beta}$ is the estimated coefficient of the exposure variable CAT.

When fitting this no interaction model to the data, we obtain estimates of the model coefficients that are listed here.

For this fitted model, then, the odds ratio is given by $e^{0.5978}$, which equals 1.82. Note that this odds ratio is a fixed number, which should be expected, as there are no interaction terms in the model.

In comparing the results for the no interaction model just described with those for the model containing interaction terms, we see that the estimated coefficient for any variable contained in both models is different in each model. For instance, the coefficient of CAT in the no interaction model is 0.5978, whereas the coefficient of CAT in the interaction model is $-12.6894$. Similarly, the coefficient of AGE in the no interaction model is 0.0322, whereas the coefficient of AGE in the interaction model is 0.0350.
It should not be surprising to see different values for corresponding coefficients as the two models give a different description of the underlying relationship among the variables. To decide which of these models, or maybe what other model, is more appropriate for this data, we need to use a strategy for model selection that includes carrying out tests of significance. A discussion of such a strategy is beyond the scope of this presentation but is described elsewhere (see Chaps. 6 and 7).

This presentation is now complete. We have described important special cases of the logistic model, namely, models for

<table>
<thead>
<tr>
<th>SUMMARY</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Introduction</td>
</tr>
<tr>
<td>2. Important Special Cases</td>
</tr>
<tr>
<td>3. Computing the Odds Ratio</td>
</tr>
</tbody>
</table>

- simple analysis
- interaction assessment involving two variables
- assessment of potential confounding and interaction effects of several covariates

We suggest that you review the material covered here by reading the detailed outline that follows. Then do the practice exercises and test.

All of the special cases in this presentation involved a (0, 1) exposure variable. In the next chapter, we consider how the odds ratio formula is modified for other codings of single exposures and also examine several exposure variables in the same model, controlling for potential confounders and effect modifiers.
I. **Overview** (page 45)
   A. Focus:
      - Simple analysis
      - Multiplicative interaction
      - Controlling several confounders and effect modifiers
   B. Logistic model formula when \( X = (X_1, X_2, \ldots, X_k) \):
      \[
P(X) = \frac{1}{1 + e^{-(a + \sum_{i=1}^{k} \beta_i X_i)}}.
      \]
   C. Logit form of logistic model:
      \[
      \text{logit } P(X) = a + \sum_{i=1}^{k} \beta_i X_i.
      \]
   D. General odds ratio formula:
      \[
      \text{ROR}_{X_i, X_0} = \prod_{i=1}^{k} e^{\beta_i (X_{i1} - X_{0i})}.
      \]

II. **Special case – Simple analysis** (pages 46–49)
   A. The model:
      \[
      P(X) = \frac{1}{1 + e^{-(a + \beta_1 E)}}
      \]
   B. Logit form of the model:
      \[
      \text{logit } P(X) = a + \beta_1 E
      \]
   C. Odds ratio for the model: \( \text{ROR} = \exp(\beta_1) \)
   D. Null hypothesis of no \( E, D \) effect: \( H_0: \beta_1 = 0 \).
   E. The estimated odds ratio \( \exp(\hat{\beta}) \) is computationally equal to \( ad/bc \) where \( a, b, c, \) and \( d \) are the cell frequencies within the four-fold table for simple analysis.

III. **Assessing multiplicative interaction** (pages 49–55)
   A. Definition of no interaction on a multiplicative scale: \( \text{OR}_{11} = \text{OR}_{10} \times \text{OR}_{01} \),
      where \( \text{OR}_{AB} \) denotes the odds ratio that compares a person in category \( A \) of one factor and category \( B \) of a second factor with a person in referent categories 0 of both factors, where \( A \) takes on the values 0 or 1 and \( B \) takes on the values 0 or 1.
   B. Conceptual interpretation of no interaction formula: The effect of both variables \( A \) and \( B \) acting together is the same as the combined effect of each variable acting separately.
C. Examples of no interaction and interaction on a multiplicative scale.

D. A logistic model that allows for the assessment of multiplicative interaction:

\[ \text{logit } P(X) = \alpha + \beta_1A + \beta_2B + \beta_3A \times B \]

E. The relationship of \( \beta_3 \) to the odds ratios in the no interaction formula above:

\[ \beta_3 = \ln \left( \frac{\text{OR}_{11}}{\text{OR}_{10} \times \text{OR}_{01}} \right) \]

F. The null hypothesis of no interaction in the above two factor model: \( H_0: \beta_3 = 0 \).

IV. The \( E, V, W \) model – A general model containing a (0, 1) exposure and potential confounders and effect modifiers (pages 55–64)

A. Specification of variables in the model: start with \( E, C_1, C_2, \ldots, C_p \); then specify potential confounders \( V_1, V_2, \ldots, V_{p_1} \), which are functions of the Cs, and potential interaction variables (i.e., effect modifiers) \( W_1, W_2, \ldots, W_{p_2} \), which are also functions of the Cs and go into the model as product terms with \( E \), i.e., \( E \times W_j \).

B. The \( E, V, W \) model:

\[ \text{logit } P(X) = \alpha + \beta E + \sum_{i=1}^{p_1} \gamma_i V_i + E \sum_{j=1}^{p_2} \delta_j W_j \]

C. Odds ratio formula for the \( E, V, W \) model, where \( E \) is a (0, 1) variable:

\[ \text{ROR}_{E=1 \text{ vs. } E=0} = \exp \left( \beta + \sum_{j=1}^{p_2} \delta_j W_j \right) \]

D. Odds ratio formula for \( E, V, W \) model if no interaction: \( \text{ROR} = \exp(\beta) \).

E. Examples of the \( E, V, W \) model: with interaction and without interaction
True or False (Circle T or F)

T F 1. A logistic model for a simple analysis involving a (0, 1) exposure variable is given by logit \( P(X) = \alpha + \beta E \), where \( E \) denotes the (0, 1) exposure variable.

T F 2. The odds ratio for the exposure–disease relationship in a logistic model for a simple analysis involving a (0, 1) exposure variable is given by \( \beta \), where \( \beta \) is the coefficient of the exposure variable.

T F 3. The null hypothesis of no exposure–disease effect in a logistic model for a simple analysis is given by \( H_0: \beta = 1 \), where \( \beta \) is the coefficient of the exposure variable.

T F 4. The log of the estimated coefficient of a (0, 1) exposure variable in a logistic model for simple analysis is equal to \( \ln(\hat{\beta}) \), where \( \hat{a}, \hat{b}, \hat{c}, \) and \( \hat{d} \) are the cell frequencies in the corresponding fourfold table for simple analysis.

T F 5. Given the model logit \( P(X) = \alpha + \beta E \), where \( E \) denotes a (0, 1) exposure variable, the risk for exposed persons \( (E = 1) \) is expressible as \( e^{\beta} \).

T F 6. Given the model logit \( P(X) = \alpha + \beta E \), as in Exercise 5, the odds of getting the disease for exposed persons \( (E = 1) \) is given by \( e^{\alpha + \beta} \).

T F 7. A logistic model that incorporates a multiplicative interaction effect involving two (0, 1) independent variables \( X_1 \) and \( X_2 \) is given by logit \( P(X) = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 \).

T F 8. An equation that describes "no interaction on a multiplicative scale" is given by \( \text{OR}_{11} = \text{OR}_{10}/\text{OR}_{01} \).

T F 9. Given the model logit \( P(X) = \alpha + \beta E + \gamma \text{SMK} + \delta E \times \text{SMK} \), where \( E \) is a (0, 1) exposure variable and \( \text{SMK} \) is a (0, 1) variable for smoking status, the null hypothesis for a test of no interaction on a multiplicative scale is given by \( H_0: \delta = 0 \).

T F 10. For the model in Exercise 9, the odds ratio that describes the exposure disease effect controlling for smoking is given by \( \exp(\beta + \delta) \).

T F 11. Given an exposure variable \( E \) and control variables AGE, SBP, and CHL, suppose it is of interest to fit a model that adjusts for the potential confounding effects of all three control variables considered as main effect terms and for the potential interaction effects with \( E \) of all
three control variables. Then the logit form of a model that describes this situation is given by logit \( P(X) = \alpha + \beta E + \gamma_1 \text{AGE} + \gamma_2 \text{SBP} + \gamma_3 \text{CHL} + \delta_1 \text{AGE} \times \text{SBP} + \delta_2 \text{AGE} \times \text{CHL} + \delta_3 \text{SBP} \times \text{CHL} \).

T  F 12. Given a logistic model of the form logit \( P(X) = \alpha + \beta E + \gamma_1 \text{AGE} + \gamma_2 \text{SBP} + \gamma_3 \text{CHL} \), where \( E \) is a (0, 1) exposure variable, the odds ratio for the effect of \( E \) adjusted for the confounding of AGE, CHL, and SBP is given by \( \exp(\beta) \).

T  F 13. If a logistic model contains interaction terms expressible as products of the form \( EW_j \) where \( W_j \) are potential effect modifiers, then the value of the odds ratio for the \( E, D \) relationship will be different, depending on the values specified for the \( W_j \) variables.

T  F 14. Given the model logit \( P(X) = \alpha + \beta E + \gamma_1 \text{SMK} + \gamma_2 \text{SBP} \), where \( E \) and SMK are (0, 1) variables, and SBP is continuous, then the odds ratio for estimating the effect of SMK on the disease, controlling for \( E \) and SBP is given by \( \exp(\gamma_1) \).

T  F 15. Given \( E, C_1, \) and \( C_2, \) and letting \( V_1 = C_1 = W_1 \) and \( V_2 = C_2 = W_2 \), then the corresponding logistic model is given by logit \( P(X) = \alpha + \beta E + \gamma_1 C_1 + \gamma_2 C_2 + E(\delta_1 C_1 + \delta_2 C_2) \).

T  F 16. For the model in Exercise 15, if \( C_1 = 20 \) and \( C_2 = 5 \), then the odds ratio for the \( E, D \) relationship has the form \( \exp(\beta + 20\delta_1 + 5\delta_2) \).
True or False (Circle T or F)

T F 1. Given the simple analysis model, logit \( P(X) = \phi + \psi Q \), where \( \phi \) and \( \psi \) are unknown parameters and \( Q \) is a \((0, 1)\) exposure variable, the odds ratio for describing the exposure–disease relationship is given by \( \exp(\phi) \).

T F 2. Given the model logit \( P(X) = \alpha + \beta E \), where \( E \) denotes a \((0, 1)\) exposure variable, the risk for unexposed persons \((E = 0)\) is expressible as \( 1/\exp(-\alpha) \).

T F 3. Given the model in Question 2, the odds of getting the disease for unexposed persons \((E = 0)\) is given by \( \exp(\alpha) \).

T F 4. Given the model logit \( P(X) = \phi + \psi HPT + \rho ECG + \pi HPT \times ECG \), where HPT is a \((0, 1)\) exposure variable denoting hypertension status and ECG is a \((0, 1)\) variable for electrocardiogram status, the null hypothesis for a test of no interaction on a multiplicative scale is given by \( H_0: \exp(\pi) = 1 \).

T F 5. For the model in Question 4, the odds ratio that describes the effect of HPT on disease status, controlling for ECG, is given by \( \exp(\psi + \pi ECG) \).

T F 6. Given the model logit \( P(X) = \alpha + \beta E + \phi HPT + \psi ECG \), where \( E \), HPT, and ECG are \((0, 1)\) variables, then the odds ratio for estimating the effect of ECG on the disease, controlling for \( E \) and HPT, is given by \( \exp(\psi) \).

T F 7. Given \( E \), \( C_1 \), and \( C_2 \), and letting \( V_1 = C_1 = W_1 \), \( V_2 = (C_1)^2 \), and \( V_3 = C_2 \), then the corresponding logistic model is given by logit \( P(X) = \alpha + \beta E + \gamma_1 C_1 + \gamma_2 C_1^2 + \gamma_3 C_2 + \delta E C_1 \).

T F 8. For the model in Question 7, if \( C_1 = 5 \) and \( C_2 = 20 \), then the odds ratio for the \( E, D \) relationship has the form \( \exp(\beta + 20 \delta) \).
Consider a 1-year follow-up study of bisexual males to assess the relationship of behavioral risk factors to the acquisition of HIV infection. Study subjects were all in the 20–30 age range and were enrolled if they tested HIV negative and had claimed not to have engaged in “high-risk” sexual activity for at least 3 months. The outcome variable is HIV status at 1 year, a (0, 1) variable, where a subject gets the value 1 if HIV positive and 0 if HIV negative at 1 year after start of follow-up. Four risk factors were considered: consistent and correct condom use (CON), a (0, 1) variable; having one or more sex partners in high-risk groups (PAR), also a (0, 1) variable; the number of sexual partners (NP); and the average number of sexual contacts per month (ASCM). The primary purpose of this study was to determine the effectiveness of consistent and correct condom use in preventing the acquisition of HIV infection, controlling for the other variables. Thus, the variable CON is considered the exposure variable, and the variables PAR, NP, and ASCM are potential confounders and potential effect modifiers.

9. Within the above study framework, state the logit form of a logistic model for assessing the effect of CON on HIV acquisition, controlling for each of the other three risk factors as both potential confounders and potential effect modifiers. (Note: In defining your model, only use interaction terms that are two-way products of the form $E \times W$, where $E$ is the exposure variable and $W$ is an effect modifier.)

10. Using the model in Question 9, give an expression for the odds ratio that compares an exposed person (CON = 1) with an unexposed person (CON = 0) who has the same values for PAR, NP, and ASCM.
**Answers to Practice Exercises**

1. T
2. F: OR = e^β
3. F: H₀: β = 0
4. F: e^β = ad/bc
5. F: risk for E = 1 is 1/[1 + e^{-(α+β)}]
6. T
7. T
8. F: OR₁₁ = OR₁₀ × OR₀₁
9. T
10. F: OR = exp(β + δSMK)
11. F: interaction terms should be E × AGE, E × SBP, and E × CHL
12. T
13. T
14. T
15. T
16. T
Logistic Regression
A Self-Learning Text
Kleinbaum, D.G.; Klein, M.
2010, XVIII, 702 p., Hardcover
ISBN: 978-1-4419-1741-6