# FACILITATOR/MENTOR GUIDE

analysis

Descriptive analysis

hypotheses

methods

design

justify

variables

Measures of association

stratify

confounding

statistical

software

Data cleaning

assess

calculate

analytic

table

shells

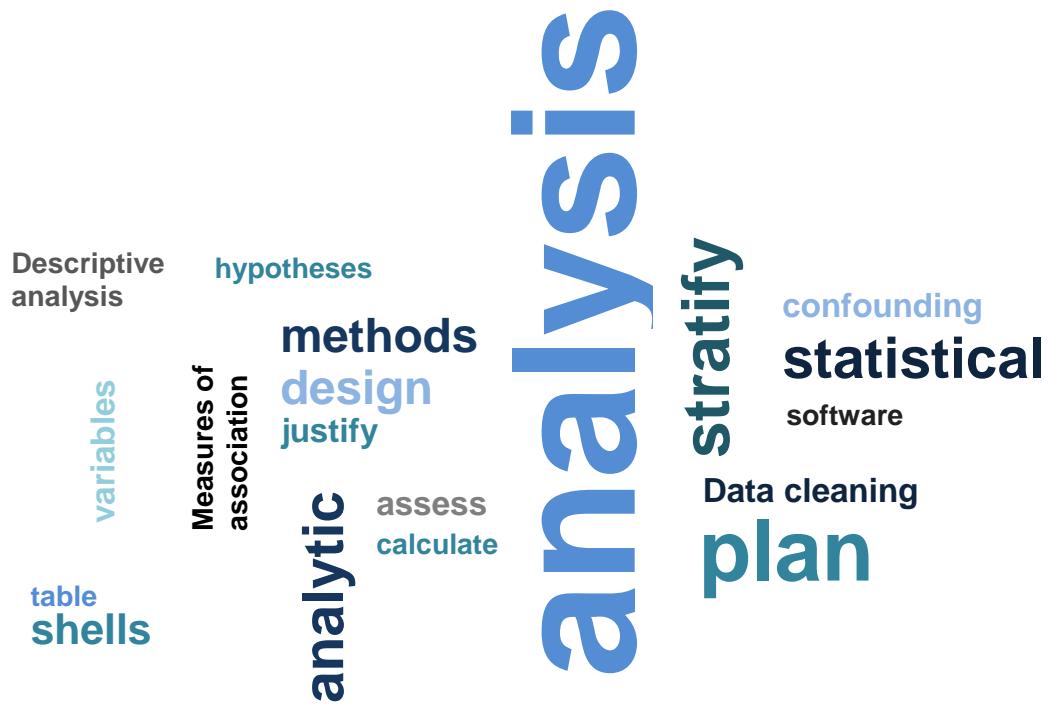plan

# Analyzing and Interpreting Large Datasets

Created: 2013

*Analyzing and Interpreting Large Datasets*. Atlanta, GA: Centers for Disease Control and Prevention (CDC), 2013.

# Table Of Contents

# Introduction

## LEARNING OBJECTIVES

At the end of this module, participants will be able to:

- Conduct and interpret descriptive analysis and analytic epidemiology
- Summarize your findings
- Identify the sections and tables to include in an analysis report

## ESTIMATED COMPLETION TIME

The module should take approximately 18 hours to complete.

## TARGET AUDIENCE

The module is designed for Field Epidemiology Training Program (FETP) residents who specialize in NCDs; however, participants can also complete the module if they are working in infectious disease.

## PRE-WORK AND PREREQUISITES

Before participating in this training module, participants must complete training in:

- Basic epidemiology and surveillance

- Basic analysis

- Statistical software program (your country is using)

- Creating an analysis plan

- Managing data (creating a data dictionary and cleaning data)

## OPTIONS FOR FACILITATING THIS TRAINING

There are two options for facilitating this training:

1. Individual mentor-directed: A mentor helps the participant complete the training. The mentor's main responsibility will be to review the mentee's work and provide feedback.

   A mentor will meet with the participant a minimum of three times. At the first meeting, the mentor should orient the participant to the training, provide examples and direction indicated, answer questions, and set future modes of contact and meeting time(s). Very small groups (less than 5 individuals) may choose to work on the training together and find individual or collective mentor(s).

2. Classroom: There are two options for classroom training. For option a, participants read the training material *prior* to attending class and then review what they read in class. For option b, participants read the training material during class.
   a. **Participants read training material *prior* to attending class.** At the start of each module section, the facilitator reviews key points. The facilitator may prepare PowerPoint slides for a brief presentation of key points, lead an informal discussion about the reading, or ask participants to answer questions individually or in small groups about what they read. (Appendix A contains sample questions.) After each review, participants will complete practice exercises and skill assessments as directed.
   b. **Participants read training material during class.** The facilitator directs students to read the training material and complete the exercises as indicated in the workbook. The facilitator leads group discussions to review what participants have read and reviews participants' answers to the exercises and skill assessments.

## MATERIALS

**For the Facilitator or Mentor:**
- Facilitator / Mentor Guide
- Flip chart and markers

**For the Participant:**
- Participant Workbook
- Supporting materials for Practice Exercises (e.g., questionnaires)
- Background information and datasets for Skill Assessment
- Laptops with statistical software

## ICON GLOSSARY

The following icons are used in this guide:

| Image Type | Image Meaning |
|---|---|
|  Activity Icon | **Pencil** - an activity, exercise, assessment or case study that participants complete |
|  Stop Icon | **Stop** - a point at which you should consult a mentor or wait for the facilitator for further locally relevant information about the topic |

| Image Type | Image Meaning |
|---|---|
| <br>Tip Icon | **Tip –** key idea to note and remember |
| <br>Resource Icon | **Resource / Website Icon**- a resource or website that may provide further information on a given topic |

## ACKNOWLEDGEMENTS

# How to Facilitate This Module

## FACILITATOR/MENTOR RESPONSIBILITIES

This training module is self-paced.  Participants learn the content by reading their workbook and participating in group discussions.  They apply what they learn by completing practice exercises and skill assessments.  Participants use a sample surveillance system evaluation for the practice exercises and one of the skill assessments.  For the other skill assessments, participants use information about an NCD surveillance system from their own country.

As a *facilitator*, you will *facilitate* or assist in the participants' learning.  Your main roles will be as follows:
- **Introduce** the module topic
- **Lead** group discussions to review or elaborate on what participants read
- **Answe**r questions that participants may have during the training
- **Review** participants' work and provide feedback
- **Be a timekeeper**, ensuring participants stay within a general schedule

General note:  As a facilitator, it will be importat to use some country-based examples as often as you can.  Subsequently, you will need to allow extra time for these practical examples to be delivered and discussed.

As a *mentor,* you will perform the same tasks and play a more active role in supporting the learner *after* the training with his or her field work.

**Note that this module has 15 brief activities (in addition to the Practice Exercises and final Skill Assessment).  Participants are asked to work with a colleague to complete these activities and check their answers with those found in Appendix A of the participant workbook.**

## OVERVIEW OF MODULE, SECTION 1: INTRODUCTION AND SECTION 2: OVERVIEW

**Total Estimated Time**: 15 - 40 minutes
**Introduction and brief overview**: 5 minutes
**Readings**: up to 25 minutes
**Group Discussion**: 10 minutes

| What To Do/What To Say |
| --- |

**Introduction and brief overview (5 minutes)**

- **Introduce** yourself if you are a new facilitator

- **Ask** participants about their experience analyzing and interpreting data and preparing reports on findings

- **Provide** an overview of what they will learn in the module

**Readings** *(Skip this step if participants have read the material prior to class.) (For those who have read the material before class, you should ask if they have any questions about the material and then discuss). This should be repeated each time the participants have read the material before the class.*

Participants read sections 1 and 2 (6 ½ pages) and complete one brief activity.

**Group discussion of what was read during or prior to class (10 minutes)**

You may use the sample questions in Appendix A as a guideline for the discussion. Participants can answer questions orally or you can provide them with the written questions and ask them to record their answers individually or in a group. Participants can use the "Key Points to Remember" section in their participant workbook to record notes or answers.

## SECTION 3: TASK B. DESCRIPTIVE ANALYSIS

**Total estimated time:** 3 -3 ½ minutes

**Readings**: up to 30 minutes

**Group Discussion**: 15 minutes

**Practice Exercise #1:** 80 minutes (including a 20 minute review)

**Readings**: up to 20 minutes

**Group Discussion**: 15 minutes

**Practice Exercise #1:** 50 minutes (including a 20 minute review)

**What To Do/What To Say**

**Readings ((*Skip this step if participants have read the material prior to class.)***

Participants read the first part of section 3 (until the stop sign, before Practice Exercise #1) and complete 2 brief activities.

**Group discussion of what was read during or prior to class (15 minutes)**

You may use the sample questions in Appendix A as a guideline for the discussion. Participants can answer questions orally or you can provide them with the written questions and ask them to record their answers individually or in a group. Participants they can use the "Key Points to Remember" section in their participant workbook to record notes or answers.

Provide any specific instructions participants will need to know about using their statistical software to:

- Run frequencies

- Create intervals

If time permits, review answers to module activities.

**Facilitate Practice Exercise #1 (1 hour plus a 20-minute review of responses)**

- **Prior to the class** be sure to run and test the variables to ensure the numbers line up correctly as shown in the answer section of this exercise

- **Divide** participants into small groups or pairs

## What To Do/What To Say

- **Ask** them to spend approximately **1 hour** completing the exercise. **Tell participants that if they finish early, they can complete the optional question (#3)**

- Make sure you are available to **answer any questions** during the exercise

- **Review** each group or pair's answers by asking participants to present their responses to the entire class or to another small group or pair for feedback

*Possible answers (in italics)*

1. Assess the variables in the tables below using descriptive statistics (e.g., mean, median, standard deviation, variance, minimum, maximum, etc).  Consider assessing variables graphically (e.g., histogram, scatterplot, etc)

| Variable: | Age (years) |
|---|---|
| Frequency | |
| Mean | |
| Median | |
| Standard deviation | |
| Minimum | |
| Maximum | |

| Variable: | Systolic blood pressure (mmHg) (1st measure) |
|---|---|
| Frequency | |
| Mean | |
| Median | |
| Standard deviation | |
| Minimum | |
| Maximum | |

## What To Do/What To Say

| Variable: | Body Mass Index ( kg/m2) |
|---|---|
| Frequency | |
| Mean | |
| Median | |
| Standard deviation | |
| Minimum | |
| Maximum | |

2. The hypertension dataset was derived using a complex design and the data are nationally representative of the civilian population in Country X.  Sample weights and sample design variables are frequently needed when analyzing data from a complex design survey.  Compare crude (i.e., unweighted) and weighted estimates.  **Examine the crude (i.e., unweighted) and weighted estimates for variables in the table below and fill in the answers.**

| | Unweighted estimate | Standard Deviation | Weighted estimate (95% CI) | Standard Error |
|---|---|---|---|---|
| **Age (mean)** | | | | |
| **Male (%)** | | | | |
| **Non-Hispanic white** | | | | |
| **Systolic blood pressure (mmHg)** | | | | |
| **Body Mass Index (kg/m2) (mean)** | | | | |
| **Hypertension (%)** | | | | |

**Optional Question:**

**What To Do/What To Say**

3. After you have explored the data, you can set up the first table using adjusted data.  It is important to provide an adequate description of your sample and include relevant health and health outcome variables.  Consider what variables would be presented in a descriptive table in a manuscript. (*Note*: Review questionnaire for available variables).

What variables would you include in the table below?  After you have selected the variables, perform the descriptive analysis and add the information to the table

| | N* | Percent | Standard Error |
|---|---|---|---|
| Sex | | | |
| Male | 467 | 45.4 | 1.3 |
| Female | 524 | 54.6 | 1.3 |
| Age | | | |
| ≤34 | 277 | 29.7 | 1.4 |
| 35-54 | 318 | 36.1 | 2.3 |
| 55-64 | 150 | 15.9 | 1.7 |
| ≥65 | 247 | 18.3 | 1.8 |
| Education | | | |
| <High School | 270 | 17.8 | 1.8 |
| HS graduate | 192 | 20.7 | 1.0 |
| Some college | 225 | 24.8 | 1.3 |
| College graduate | 184 | 26.4 | 1.1 |
| Under 25 | 118 | 10.2 | 1.0 |
| Race/ethnicity | | | |
| Non-Hispanic white | 478 | 68.9 | 3.2 |
| Non-Hispanic black | 175 | 10.9 | 1.8 |
| Hispanic | 285 | 13.3 | 2.9 |
| Other | 55 | 6.9 | 1.5 |
| Body Mass Index | | | |
| Underweight/Normal | 275 | 29.8 | 1.6 |
| Overweight | 317 | 32.2 | 1.5 |
| Obese | 390 | 38.0 | 1.4 |

*Unweighted

**Readings *(Skip this step if participants have read the material***

| What To Do/What To Say |
|---|

*prior to class.)*

Participants read 3 ½ pages about bivariable analysis and complete 5 brief activities.

**Group discussion of what was read during or prior to class (15 minutes)**

You may use the sample questions in Appendix A as a guideline for the discussion. Participants can answer questions orally or you can provide them with the written questions and ask them to record their answers individually or in a group.  Participants they can use the "Key Points to Remember" section in their participant workbook to record notes or answers.

**Provide any specific instructions participants will need to know about using their statistical software to do cross-tabs.**

If time permits, review answers to module activities.

**Facilitate Practice Exercise #2 (50 minutes plus a 20-minute review of responses)**
- **Keep** participants in their same small groups or pairs
- **Ask** them to spend approximately **30 minutes** completing the exercise
- **NOTE: Point out that the tables show unweighted data and explain why**
- Make sure you are available to **answer any questions** during the exercise
- **Review** each group or pair's answers by asking participants to present their answers to the entire class or to another small group or pair for feedback

*Possible answers (in italics):*

1.  How would you compare your health outcome of interest (hypertension) by descriptive characteristics to assess for patterns in the data?

**Hypertension by Sex**

| | Hypertension | | | | | |
|---|---|---|---|---|---|---|
| **Sex** | Yes | | | No | | |
| | N* | % | 95% CI | N* | % | 95% CI |
| Male | | | | | | |
| Female | | | | | | |

**Hypertension by Racial/Ethnic Group**

| | Hypertension | | | | | |
|---|---|---|---|---|---|---|
| **Race** | Yes | | | No | | |
| | N* | % | 95% CI | N* | % | 95% CI |
| Non-Hispanic White | | | | | | |
| Non-Hispanic Black | | | | | | |
| Hispanic | | | | | | |

**Hypertension by Age Group**
**(Alternate table view)**

| | Hypertension | | |
|---|---|---|---|
| Age (group) | Yes | | |
| | N* | % | 95% CI |

| | | | |
|---|---|---|---|
| ≤34 years | | | |
| 35-54 years | | | |
| 55-64 years | | | |
| >65 years | | | |

**\*Unweighted N**

## SECTION 4: ANALYTIC EPIDEMIOLOGY

**Total estimated time:** 2 ¾ - 4 hours
    **Readings**: up to 35 minutes
    **Group discussion**: 15 minutes
    **Practice Exercise #3**: 80 minutes (including a 20-minute review)
    **Readings:** up to 35 minutes
    **Group discussion:** 15 minutes
    **Practice Exercise #4:** 65 minutes (including a 20-minute review)

---

### What To Do/What To Say

**Readings** *(Skip this step if participant have read the material prior to class.)*

Participants read approximately 6 ½ pages and complete 2 brief activities.

**Group discussion of what was read during or prior to class (15 minutes)**
You may use the sample questions in Appendix A as a guideline for the discussion. Participants can answer questions orally or you can provide them with the written questions and ask them to record their answers individually or in a group.  Participants they can use the "Key Points to Remember" section in their participant workbook to record notes or answers.

Provide any specific instructions participants will need to know about using their statistical software to:
Compute PR and POR
Perform statistical testing.

If time permits, review answers to module activities.

**Facilitate Practice Exercise #3 (1 hour plus a 20-minute review of responses)**
- **Keep** participants in the same small groups or pairs
- **Ask** them to spend approximately **1 hour** completing the exercise
- Make sure you are available to **answer any questions** during the exercise

## What To Do/What To Say

- **Review** each group or pair's answers by asking participants to present their answers to the entire class or to another small group or pair for feedback.

*Possible answers (in italics):*

1. How would you additionally assess associations between hypertension and descriptive characteristics? (Consider: Is hypertension more frequent in male compared to females?) You may wish to create additional derived variables for these analyses to simplify the associations. (*Note:* Statistical significance testing is included the in next exercise).

### Hypertension by <u>*Sex*</u>

| | Hypertension | | | | | |
|---|---|---|---|---|---|---|
| *Sex* | Yes | | | No | | |
| | N* | % | 95% CI | N* | % | 95% CI |
| *Male* | *176* | *36.4* | *30.1-43.1* | *291* | *63.6* | *56.9-69.9* |
| *Female* | *186* | *30.8* | *23.4-39.3* | *337* | *69.2* | *60.7-76.6* |

**\*Unweighted N**
PR = <u>*1.143 (95% CI: 0.869-1.898)*</u>
POR = <u>*1.28 (95% CI: 0.869-1.898)*</u>

Note: The OR is often rounded to one decimal and the CI rounded to two decimals (e.g., POR = 1.3 ((0.87-1.90). For the purpose of this training, we have left in the full numbers and not rounded the answers.

**What To Do/What To Say**

**Hypertension by _Age Group_**

| Age group | Hypertension | | | | | |
|---|---|---|---|---|---|---|
| | Yes | | | No | | |
| | N* | % | 95% CI | N* | % | 95% CI |
| <55 years | 95 | 16.7 | 14.6-19.2 | 500 | 83.3 | 80.8-85.4 |
| ≥ 55 years | 267 | 65.2 | 56.7-72.7 | 129 | 34.8 | 27.3-43.3 |

**\*Unweighted N**
PR = _0.257 (95% CI: 0.231-0.286)_
POR = _0.108 (95% CI: 0.082-0.141)_

2. Interpret your findings.  For example, if the prevalence of hypertension is greater in females than males, how would you describe your findings?

Interpretation:
_In the Hypertension by Sex example above, the PR is 1.14 and the POR is 1.28.  PR: The probability having hypertension is 1.1 times more likely in males than females.  POR: Odds of hypertension is 1.3 times higher in males compared with females._

**Readings _(Skip this step if participants have read the material prior to class.)_**

Participants read approximately 10 pages and complete 2 brief activity.

**Group discussion of what was read during or prior to class (15 minutes)**

You may use the sample questions in Appendix A as a guideline for the discussion. Participants can answer questions orally or you can provide them with the written questions and ask them to record their answers individually or in a group.  Participants they can use the "Key Points to Remember" section in their participant workbook to record notes or answers.

## What To Do/What To Say

**Provide any specific instructions participants will need to know about using their statistical software to do stratified analysis and crude analysis.**

If time permits, review answers to the module activity.  If time is short, provide the correct answers so they can verify later.

**Facilitate Practice Exercise #4(45 minutes plus a 20-minute review of responses)**
- **Keep** participants in the same small groups or pairs
- **Ask** them to spend approximately **45 minutes** completing the exercise
- Make sure you are available to **answer any questions** during the exercise
- **Review** each group or pair's answers by asking participants to present their answers to the entire class or to another small group or pair for feedback

Possible answers (*in italics*):

1. What is the first step in assessing the relationship between obesity and hypertension?

   *Assess the relationship between non-obese/obese and hypertension.*

2. Fill in the table below.

**Hypertension by _Obesity Classification_**

|  | Hypertension | | | | | |
|---|---|---|---|---|---|---|
| *Obese* | Yes | | | No | | |
|  | N* | % | 95% CI | N* | % | 95% CI |
| *Obese* | *189* | *47.6* | *40.5-54.8* | *200* | *52.4* | *45.2-59.5* |
| *Non-Obese* | *170* | *24.6* | *19.5-30.5* | *422* | *75.4* | *69.5-80.5* |

**\*Unweighted N**

## What To Do/What To Say

PR = *1.933 (95% CI: 1.619-2.308)*
POR = *2.78 (95% CI: 2.146-3.600)*
$\chi^2$ = *54.798, df= 1, p= <0.01*

3. Your findings indicate the relationship between obesity and hypertension. Are there variables that have been collected in the survey that may distort the relationship between obesity and hypertension? Would you expect gender to confound the relationship? Complete the tables below.

**Hypertension by *Weight Classification***
**Gender: Males**

| | Hypertension | | | | | |
|---|---|---|---|---|---|---|
| *Weight Classification* | Yes | | | No | | |
| | N* | % | 95% CI | N* | % | 95% CI |
| *Obese* | 80 | 46.0 | 37.2-55.0 | 90 | 54.0 | 45.0-62.8 |
| *Non-Obese* | 93 | 30.1 | 23.2-37.9 | 198 | 30.1 | 23.2-37.9 |

**\*Unweighted N**
PR = *1.529 (95% CI: 1.151-2.033)*
POR = *1.98 (95% CI: 1.255-3.124)*

**Hypertension by *Weight Classification***
**Gender: *Females***

| | Hypertension | | | | | |
|---|---|---|---|---|---|---|
| *Weight Classification* | Yes | | | No | | |
| | N* | % | 95% CI | N* | % | 95% CI |
| Obese | 109 | 49.0 | 40.1-57.9 | 110 | 51.0 | 42.1-59.9 |
| Non-Obese | 76 | 20.1 | 13.3-29.1 | 223 | 79.9 | 70.9-86.7 |

**What To Do/What To Say**

**\*Unweighted N**
PR = *2.441 (95% CI: 1.734-3.438)*
POR = *3.825 (95% CI: 2.455-5.960*)

4. Are there differences among the demographic variables that you assessed?  How do you interpret the findings?

   *Proportion and odds of hypertension still higher among obese females and males after stratification.  The PR and POR are higher among females and males overall, with wide 95% CIs.*

***Note: Additional example of a method of confounding exploration:***
***Is gender a risk factor for hypertension? (Among non-obese)***

| | *Hypertension* | | |
|---|---|---|---|
| *Gender* | *Yes* | *No* | *Total* |
| *Male* | | | |
| *Female* | | | |
| *Total* | | | |

***Is obesity associated with gender?***

| | *Obesity* | | |
|---|---|---|---|
| *Gender* | *Yes* | *No* | *Total* |
| *Male* | | | |
| *Female* | | | |
| *Total* | | | |

*Last question – is gender in the pathway between obesity and hypertension? No*

## SECTION 5: INTERPRETING AND REPORTING YOUR FINDINGS

**Total Estimated Time**: 75 – 95 minutes
**Readings**: up to 20 minutes
**Group Discussion**: 15 minutes
**Practice Exercise** #5: 60 minutes (includes a 20 minute review)

| What To Do/What To Say |
| --- |

**Readings** *(Skip this step if participants have read the material prior to class.)*
Participants read approximately 3 ½ pages and complete 3 brief activities.

**Group discussion of what was read during or prior to class (15 minutes)**
Review the answers to the activities. Briefly discuss ways to present findings to stakeholders, including components of a 2 to 3 page report.

**Facilitate Practice Exercise #5 (45 minutes plus a 20-minute review of responses)**
- **Keep** participants in the same small groups or pairs
- **Ask** them to spend approximately **45 minutes** completing the exercise
- Make sure you are available to **answer any questions** during the exercise
- **Ask each group to present its findings to the rest of the class**

Possible answers (*in italics*):

1. Based on the results of your analyses, use the space below to summarize your key findings.

    *Review the research questions. What was the prevalence of hypertension in the population? Where there differences in hypertension prevalence among different demographic groups? Was obesity related to hypertension?*

    *Other findings to include:*

## What To Do/What To Say

*Demographic and descriptive differences, including percent, 95% confidence intervals*
*Prevalence ratio and Prevalence odds ratio, and interpretation*
*Summary of findings when assessing confounding or effect modification*

2. What main sections would you include in the report?  List in the space below.

   - *Background of problem*

   - *Describe methods*

   - *Show results (main findings)*

   - *Conclusions*

   - *Recommendations*

3. Which of the tables you created would you include in the report to support your findings?  Describe them in the space below.

   *A variety of tables could be included.  A demographic table describing the survey sample would be useful for an initial table. Secondary tables may include significant comparisons among demographic groups, including the measures of association and statistical testing.*

4. Would you recommend changes to the national survey to better assess hypertension in Country X?

   *Additional variables would have been useful in this analysis.  For example, physical activity and dietary intake are significantly related to hypertension, but were not included.  Additional sociocultural and/or psychosocial variables may have also been useful – stress, social support, depression, etc.*

## SKILL ASSESSMENT



**Activity**

**Estimated Time:** 4 hours, plus 1 hour to present findings.

Participants will practice conducting descriptive analysis and analytic epidemiology.

**Note: Because everyone's dataset will be different, it is difficult to create one skill assessment with questions to meet everyone's needs. Therefore, no Activity Workbook will accompany this module. You will need to refer to the Appendix and adapt the questions and tables accordingly.**

**Distribute** the background information participants will need to complete the skill assessment for a study in their own country, such as:
- Sample questionnaires
- Background information about the type of analysis they will need to conduct and instructions. (See Appendix for sample questions and tables you will need to modify.)

**Remind** participants to use the data dictionary and table shells they created in the previous two modules.

**Observe** participants completing the assignment and be able to answer any questions.

**Ask** participants to share their findings with the rest of the class. (Allow for 10 minute presentations.)

## CONCLUSION

**Estimated Time:** 20 minutes

**Ask** participants to provide some main points that they learned in the module.

**Ask** participants for their reactions to what they learned in the training and how they will apply the skills.

# Appendices

**Sample Review Questions for Sections 1 -2: Introduction and Overview**

1. What types of questions does descriptive epidemiology and analytic epidemiology help answer?
   *Possible answer:*
   *Descriptive epidemiology helps to answer the what, who, where and when.  Analytic epidemiology helps to answer the why and the how.*

2. What are some commonly used datasets used for NCDs?
   *Possible answer:*
   *Vital registration, DHS, WHO STEPS survey, NHANES, BRFSS*

3. Why do you need to show standard error in a descriptive analysis table, for example, a table that shows 53.1% men and 46.9% women in a sample, with an SE of .87
   *Possible answer:*
   *To show how much the mean of the sample differs from the mean of the entire population. The smaller the SE the closer the mean of the sample to the true population*

4. A descriptive analysis table shows the mean age of 48.8 for an unweighted estimate and a standard deviation of 18.7.  What does the standard deviation of 18.7 mean?
   *Possible answer:*
   *It reflects the variability of the distribution of the age variable.*

5. What is a weighted sample and why is it important?
   *Possible answer:*
   *A weighted sample is representative of the population.  Weights are used to account for complex survey design (including oversampling), survey non-response, and post-stratification.*

**Sample Review Questions for Sections 3: Descriptive Analysis (Univariable Analysis)**

1. Give an example of when you conducted descriptive analysis.  What type of information did you assess for continuous variables?  What type of information did you assess for discrete variables?

*Answers should include: analyzing assessing information such as range, mean, median and mode for continuous variables and the range and frequency of discrete variables.*

2. Why would you run a frequency?

   *Possible answers: to provide an organized picture of the data to see how individual scores are distributed on a specified scale of measurement.*

3. Provide an example of when you had to create variables during descriptive analysis.

   *Answers should include: when you want to group continuous variable data into logical intervals or categories.*

4. True or False: If your analysis shows a very small number of responses, such as "don't know", eliminating the information is an acceptable choice

   *Answer: true*

**Sample Review Questions for Sections 3: Descriptive Analysis (Bivariable Analysis)**

1. Give an example of when you conducted bivariable analysis. What type of patterns or connections between demographic and / or exposure characteristics did you make?

   *Answers should include: establishing similarities or differences of demographic characteristics (such as age and gender) and/or exposure characteristics (such as environmental exposure and diet).*

2. What are cross-tabs?

   *Possible answers: A cross-tab or cross-tabulation is a two or more dimensional table that record the number of respondents that have the specific characteristics described in the cells of the table.*

3. When would you use cross-tabs?
   *Possible answers:*
   - *To look at relationships among two or three variables*

- *When you want a descriptive statistical measure to determine whether differences among groups are large enough to indicate some sort of relationship among variables.*

4. True or False:  You can use cross-tabs to test a hypothesis about the relationship between two or more variables.

   *Answer:  False.  Use analytic epidemiology.*

## Sample Review Questions for Section 4: Analytic Epidemiology (Up to  Exercise #3)

1. What does a PR of 1.26 in table 7 indicate?

   *Answer: The prevalence of coronary heart disease (CHD) is 1.26 times as high (e.g., 26% higher) among persons who consume more than 3 alcoholic drinks per day than among persons who consume 3 or fewer drinks per day.*

2. What does a POR of 1.6 in table 8 indicate?

   *Answer:  The odds of having CHD for people who drink more than 3 drinks a day is 1. 6 as great as the odds of having CHD for people who drink less than or equal to 3 drinks a day.*

3. What does a very small p-value indicate when you conduct statistical tests?

   *Answer: A very small p-value means that you would unlikely observe such an association if the null hypothesis were true.*

4.  What does a wide CI indicate?

   *Answer: A wide confidence interval (e.g., 1.71 – 5.42), indicates a large amount of variability or imprecision.*

5. What does a narrow CI indicate?

   *Answer: A narrow confidence interval (e.g., 1.62 – 2.2), indicates little variability and high precision.*

**Sample Review Questions for Section 4: Analytic Epidemiology (Stratified Analysis, EMM, Confounding)**

1.  How can you assess effect measure modification?

    *Answer:  By stratifying the analysis by a third variable and determining if the PORs are different from each other OR*
    *Looking at the confidence intervals around each stratum-specific measure.  If the CIs overlap, no EMM is present.*

2.  When is EMM present?

    *Answers should include:*
    *EMM is present when the stratum-specific measures of association are different from each other.*

3.  How can you assess for confounding?

    *Answers should include:  Compare the crude measure of association to the stratified measures of association.*

4.  How would you determine if the covariate is the confounder? What action should you take?

    *Answer: If the crude measure and the stratified measures are **close in value**, the covariate has **no impact** on the exposure-outcome relationship.  Report the crude measure.*

5.  What does it mean if the crude measure and the stratified measures are different? What action should you take?

    *Answer: If the crude measure and the stratified measures are **different**, the covariate is a **confounder**.  Take steps to control for confounding.*

## APPENDIX B

**Skill Assessment**

*Facilitator note:  Modify the following tables and questions accordingly.*

**Part 1: Conduct Descriptive Statistics**

1.  Assess the variables in the tables below using descriptive statistics (e.g., frequency, mean, median, standard deviation, and minimum, maximum).  Consider assessing variables graphically (e.g., histogram, scatterplot, etc).

| Variable: | |
|---|---|
| Frequency | |
| Mean | |
| Median | |
| Standard deviation | |
| Minimum | |
| Maximum | |

| Variable: | |
|---|---|
| Frequency | |
| Mean | |
| Median | |
| Standard deviation | |
| Minimum | |
| Maximum | |

| Variable: | |
|---|---|
| Frequency | |
| Mean | |
| Median | |
| Standard deviation | |
| Minimum | |
| Maximum | |

Examine the crude (i.e., unweighted) and weighted estimates for variables in the table below and fill in the answers.

|  | **Unweighted estimate** | **Standard Deviation** | **Weighted estimate (95% CI)** | **Standard Error** |
|---|---|---|---|---|
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

2. After you have explored the data, you can set up the first table using adjusted data. It is important to provide an adequate description of your sample and include relevant health and health outcome variables. Consider what variables would be presented in a descriptive table in a manuscript. (*Note*: Review questionnaire for available variables).

What variables would you include in the table below? After you have selected the variables, perform the descriptive analysis and add the information to the table.

|  | **N** | **Percent** | **Standard Error** |
|---|---|---|---|
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

## Part 2: Conduct Bivariable Analysis

1. How would you compare your health outcome of interest by descriptive characteristics to assess for patterns in the data? **Use the tables you created in the Analysis Plan module, such as the ones below:**

_____ **by Sex**

| Sex | Yes | | | No | | |
|-----|-----|-----|--------|------|-----|--------|
| | N* | % | 95% CI | N* | % | 95% CI |
| Male | | | | | | |
| Female | | | | | | |

***Unweighted N**

_____ **by Racial/Ethnic Group**

| Race | Yes | | | No | | |
|------|-----|-----|--------|------|-----|--------|
| | N* | % | 95% CI | N* | % | 95% CI |
| Non-Hispanic White | | | | | | |
| Non-Hispanic Black | | | | | | |
| Hispanic | | | | | | |

***Unweighted N**

**Part 3: Concepts of Association**

1. How would you additionally assess associations between _____ and descriptive characteristics? (Consider: Is _____ more frequent in male compared to females?) You may wish to create additional derived variables for these analyses to simplify the associations. **Use the tables you created in the Analysis Plan module, similar to the ones below.** Select the appropriate statistical test, specify the p value, and perform the statistical test.

| **Exposure Variable:** _____ | **Outcome Variable:** | | | | | |
|---|---|---|---|---|---|---|
| | Yes | | | No | | |
| | N* | % | 95% CI | N* | % | 95% CI |
| | | | | | | |
| | | | | | | |

**\*Unweighted N**

PR =_____ (95% CI:_____-_____)
POR = _____(95% CI: _____-_____)
$\chi^2$ =_____ , df=_____, p=_____

| **Exposure Variable:** _____ | **Outcome Variable:** | | | | | |
|---|---|---|---|---|---|---|
| | Yes | | | No | | |
| | N* | % | 95% CI | N* | % | 95% CI |
| | | | | | | |
| | | | | | | |

**\*Unweighted N**

PR =_____ (95% CI:_____-_____)
POR = _____(95% CI: _____-_____)
$\chi^2$ =_____ , df=_____, p=_____

| Exposure Variable: _____ | Outcome Variable: | | | | | |
|---|---|---|---|---|---|---|
| | Yes | | | No | | |
| | N* | % | 95% CI | N* | % | 95% CI |
| | | | | | | |
| | | | | | | |

**\*Unweighted N**

PR =_____ (95% CI:_____-_____)

POR = _____(95% CI: _____-_____)

$\chi^2$ =_____ , df=_____, p=_____

| Exposure Variable: _____ | Outcome Variable: | | | | | |
|---|---|---|---|---|---|---|
| | Yes | | | No | | |
| | N* | % | 95% CI | N* | % | 95% CI |
| | | | | | | |
| | | | | | | |

**\*Unweighted N**

PR =_____ (95% CI:_____-_____)

POR = _____(95% CI: _____-_____)

$\chi^2$ =_____ , df=_____, p=_____

2.  Interpret your findings and record in the space below.

**Part 4: Assess for Confounding and EMM**

1.  **Use the table shells created in the Analysis Plan module** to assess for confounding and EMM.  Recreate the tables and fill them in below.

|  |  |  |  |  |  |  |
|---|---|---|---|---|---|---|
|  | Yes |  |  | No |  |  |
|  | N* | % | 95% CI | N* | % | 95% CI |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |

**\*Unweighted N**
  PR =
  POR =

|  |  |  |  |  |  |  |
|---|---|---|---|---|---|---|
|  | Yes |  |  | No |  |  |
|  | N* | % | 95% CI | N* | % | 95% CI |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |

**\*Unweighted N**
  PR =
  POR =

## Part 5: Interpreting and Reporting Findings

1. Based on the results of your analyses, use the space below to summarize your key findings.

2. What main sections would you include in the report?  List in the space below.

3. Which of the tables you created would you include in the report to support your findings?  Describe them in the space below.