

FACILITATOR/MENTOR GUIDE



Creating an Analysis Plan

Created: 2013



Creating an Analysis Plan. Atlanta, GA: Centers for Disease Control and Prevention (CDC), 2013.

Table of Contents

INTRODUCTION	3
LEARNING OBJECTIVES.....	3
ESTIMATED COMPLETION TIME	3
TARGET AUDIENCE.....	3
PRE-WORK AND PREREQUISITES.....	3
OPTIONS FOR FACILITATING THIS TRAINING	3
MATERIALS	4
CONFIGURATION OF THE TRAINING ROOM	5
ACKNOWLEDGEMENTS.....	6
HOW TO FACILITATE THIS MODULE	8
FACILITATOR / MENTOR RESPONSIBILITIES.....	8
CLASS INTRODUCTIONS, OVERVIEW OF WORKSHOP AND SECTIONS 1, 2 AND 3	9
SECTION 3: PREPARING TABLE SHELLS	12
SKILL ASSESSMENT	20
CONCLUSION	21
APPENDIX A.....	22
APPENDIX B.....	23

Introduction

LEARNING OBJECTIVES

Given information about a noncommunicable (NCD) health problem and a request for health-related information, participants will be able to create an analysis plan that includes the following:

- Research question(s) and/or hypotheses,
- Dataset(s) to be used,
- Inclusion/exclusion criteria,
- Variables to be used in the main analysis,
- Statistical methods and software to be used, and,
- Table shells to prepare for:
 - Univariable analysis,
 - Bivariable analysis,
 - Calculating measures of association, and,
 - Assessing for confounding and effect measure modification.

ESTIMATED COMPLETION TIME

The workbook should take between 6 and 7 hours to complete.

TARGET AUDIENCE

The workbook is designed for FETP fellows who specialize in NCDs; however, participants can also complete the module if they are working in the infectious diseases area.

PRE-WORK AND PREREQUISITES

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

- Basic epidemiology and surveillance
- Basic data analysis concepts

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 two 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 B contains sample questions.) After each review participants will complete practice exercises and skill assessments as directed.
 - b. **Participants read training material during class¹:** 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
- Activity Workbook
- Supporting materials for Practice Exercises (e.g., questionnaires)
- Background information for Skill Assessment

¹ See Appendix A for sample class agenda for the three data analysis modules.






CONFIGURATION OF THE TRAINING ROOM

If this training will be implemented in a facilitator-led setting, please note the following recommendations:

1. Use a room large enough to host breakout groups of 6–8 participants.
2. Each breakout group should have one rectangular or round table for completing small group work.
3. An ideal training room will have enough space between tables to have flip charts for each group and enough space between tables so that groups will not be too distracted by each other.

ICON GLOSSARY

The following icons are used in this guide:

Image Type	Image Meaning
 Group Icon	Group discussion that you will lead, either to review key points or answers to an activity
 Flip Chart Icon	Write responses during facilitator-led discussions or debriefs
 Activity Icon	Activity, exercise, assessment or case study that participants complete
 Prepare Icon	Activity in the module for which you need to prepare (e.g., making handouts of a report, identifying a local example)
 Reading Icon	Participants read a section in the participant workbook

PREPARATION

There are several facilitator/mentor-led discussions and activities throughout this module. Be sure to review this facilitator guide and read the descriptions of the following discussions and activities:

- Introduction to the module (page 8)
- Prepare a brief discussion of what participants have read in sections 1, 2 and half of section 3. You may use the sample questions in Appendix B.
- Prepare an example of a country-specific NCD study so that participants can create table shells (page 11).
- Prepare background information for the Skill Assessment (page 18).

ACKNOWLEDGEMENTS

Many thanks to the following colleagues from the Centers for Disease Control and Prevention for:

1) Providing detailed feedback and guidance:

- Lina Balluz, ScD, MPH, Office of Surveillance, Epidemiology and Laboratory, Division of Behavioral Surveillance
- Richard Dicker, MD, MS, Center for Global Health, Division of Global Health Protection
- Antonio Neri, MD, MPH, National Center for Chronic Disease Prevention and Health Promotion, Division of Cancer Prevention and Control
- Mona Saraiya, MD, MPH, National Center for Chronic Disease Prevention and Health Promotion, Division of Cancer Prevention and Control

2) Developing the hypertension case study for the practice exercises:

- Fleetwood Loustalot, PhD, FNP, National Center for Chronic Disease Prevention and Health Promotion, Division of Heart Disease and Stroke Prevention
- Andrea Neiman, MPH, PhD, National Center for Chronic Disease Prevention and Health Promotion, Division of Heart Disease and Stroke Prevention
- Cathleen Gillespie, MS, National Center for Chronic Disease Prevention and Health Promotion, Division of Heart Disease and Stroke Prevention
- Edward Gregg, PhD, National Center for Chronic Disease Prevention and Health Promotion, Division of Diabetes Translation

Some of the content of this module was taken from a training manual developed by the CDC, Division of Epidemiology and Surveillance Capacity Development: *Advanced Management and Analysis of Data Using Epi Info for Windows: Risk Factors for Sexually Transmitted Infections in Kuwadzana, Zimbabwe; 2006.*

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 hypertension case study for the practice exercises; for the skill assessments they use information about an NCD study 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.
- **Answer** questions that participants may have during the training.
- **Review** participants' work and provide feedback.
- **Be a timekeeper and**, ensure participants stay within a general schedule.

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.

CLASS INTRODUCTIONS, OVERVIEW OF WORKSHOP AND SECTIONS 1, 2 AND 3 (UNTIL TABLE SHELLS)

Total Estimated Time: 1 ½ - 2 ¼ hours


Workshop Introduction and Overview of Training: 30 minutes






Readings and Review Activity: up to 45 minutes

Group Discussion: 10 minutes

Practice Exercise #1: 45 minutes

Debrief: 15 minutes

Duration/Session Type	What to Do/What to Say
<div data-bbox="272 1157 380 1262" style="text-align: center;">  </div> <p data-bbox="240 1272 412 1339" style="text-align: center;">Group Discussion</p> <p data-bbox="240 1381 412 1415" style="text-align: center;">30 Minutes</p>	<p data-bbox="488 663 1230 697">Workshop Introduction and Overview of Training</p> <ul style="list-style-type: none"> <li data-bbox="500 711 1344 953">• Introduce yourself and ask participants to introduce one another by providing: <ul style="list-style-type: none"> <li data-bbox="607 795 818 829">○ Their name <li data-bbox="607 837 894 871">○ Where they work <li data-bbox="607 879 1170 913">○ Experience conducting data analysis <li data-bbox="607 921 1081 955">○ Expectations for the workshop <li data-bbox="500 963 1386 1081">• Distribute the Participant Workbook and Activity Workbook (for this module or all three modules if conducting a 5-day workshop). <li data-bbox="500 1089 1403 1375">• Explain how participants will learn the skills by reading the Creating an Analysis Plan Participant Workbook and then apply what they have read by completing two practice exercises. Facilitator-led discussions will clarify or elaborate on these exercises. Describe how the activity workbook contains skill assessments where they will apply <i>all</i> of the skills learned in the module to their own NCD study. <li data-bbox="500 1383 1422 1892">• Provide an overview of the module and the skills taught to prepare them for data analysis: <ul style="list-style-type: none"> <li data-bbox="558 1488 1422 1648">○ Creating an Analysis Plan module explains how to develop a plan for analysis, including identifying datasets and statistical software, and creating table shells to be used in the analysis. <li data-bbox="558 1677 1386 1795">○ Managing Data module explains how to create a data dictionary to use in the analysis and how to clean the data. <li data-bbox="558 1824 1373 1892">○ Analyzing and Interpreting Large Datasets module explains how to conduct descriptive and analytic

Duration/Session Type	What to Do/What to Say
	epidemiology and how to interpret and present the findings.
 <p>Activity</p>  <p>Reading</p> <p>45 minutes</p>	<p>Readings and Review Activity</p> <p>Tell participants to read sections 1 and 2, and part of section 3 until they see the “STOP” sign (page 7). They will also complete a brief review activity about types of statistical data. (Skip this step if participants have completed the reading prior to class.)</p>
 <p>Group Discussion</p> <p>10 minutes</p>	<p>Group Discussion</p> <ul style="list-style-type: none"> • Lead a brief discussion on what participants have read. • You may use the sample questions in Appendix B 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.
 <p>Activity</p> <p>45 minutes</p>	<p>Practice Exercise #1</p> <ul style="list-style-type: none"> • Divide participants into small groups or pairs. • Distribute the sample questionnaires for the hypertension case study. • Ask them to spend approximately 45 minutes completing the exercise (including reading the hypertension case study and answering the questions). • Make sure you are available to answer any questions during the exercise.
 <p>Group Discussion</p>	<p>Debrief</p> <ul style="list-style-type: none"> • Review each group or pair’s answers. If time permits, you can ask participants to present their answers to the entire class or to another small group or pair for feedback. • Possible answers:

Duration/Session Type	What to Do/What to Say
15 minutes	<p>Research question(s) and/or hypotheses:</p> <ul style="list-style-type: none"> ○ <i>What are the demographic characteristics of the population? Age, sex, education level, racial/ethnic group, geographic region?</i> ○ <i>What are the relevant behavioral factors and/or health conditions relevant to hypertension? Physical activity, diet, tobacco use, overweight/obesity?</i> ○ <i>What is the prevalence and severity of hypertension in the population?</i> ○ <i>What are the prevalence of awareness, treatment, and control of high blood pressure?</i> ○ <i>What segments of the population are at highest risk and the best candidates for preventive intervention?</i> <p>Dataset(s) to be used: <i>National health survey data (includes behavioral risk factors and biometric measurements, such as blood pressure, height, weight) and provincial hospital data</i></p> <p>Inclusion/exclusion criteria: <i>All records in the national health survey and provincial hospital data that pertain to hypertension and the risk factors (physical activity, diet, tobacco use, overweight/obese)</i></p> <p>Variables to be used in the main analysis: (List 3 or 4 outcomes and exposure variables. You will create a data dictionary in another module.) <i>The variables may be included in the national health survey which included demographic and descriptive characteristics and biometric measurements. The survey has been conducted every two years for the past decade. The most recent health survey data were collected last year. The provincial hospital data would be able to provide a more focused look at admissions and discharge data, including costs of care and services used. Detailed dietary recall data are not available.</i></p> <p><i>Note: The amount of questions in the survey has been limited because of brevity, understanding, time, etc. Other relevant risk factors would likely be included in larger surveys. This could be</i></p>

Duration/Session Type	What to Do/What to Say
	<p><i>a teaching point about survey development, or a recommendation for future survey implementation.</i></p> <p>Statistical methods and software to be used:</p> <p><i>Univariable analysis (descriptive statistics), Bivariable analysis, Multivariable analysis</i></p> <p><i>Participants should list the statistical software used in their country (e.g., SPSS, SAS).</i></p>

SECTION 3: PREPARING TABLE SHELLS





Total estimated time: 1 hour, 50 minutes – 2 ½ hours





Readings and Activities: up to 40 minutes

Group Discussion: 20 minutes

Practice Exercise #2: 1 hour

Debrief: 30 minutes

Duration/Session Type	What to Do/What to Say
 <p>Activity</p>  <p>Reading</p> <p>40 minutes</p>	<p>Readings and Activities</p> <p>Participants read 5 pages (until page 18) and complete two, brief activities.</p>
 <p>Prepare</p>  <p>Group</p>	<p>Group Discussion</p> <ul style="list-style-type: none"> • Provide an example of an NCD study that has been requested or prepared (you may create this information or use a country-specific example of a priority NCD). • Ask for volunteers to draw and record examples of: <ul style="list-style-type: none"> ○ A univariate table shell ○ A bivariable table shell

Duration/Session Type	What to Do/What to Say
<p>Discussion</p>  <p>Flip Chart</p> <p>20 minutes</p>	<ul style="list-style-type: none"> ○ Variable pairs to test for statistical associations using chi-square and t-tests as appropriate ○ A table shell to prepare for calculating measures of association ○ Potential confounders ○ Table shells to assess for confounding and effect measure modification (interaction) <ul style="list-style-type: none"> ● Include in the discussion a brief review of how to select the referent (or reference) category. Explain that you would select the category that has the lowest level of exposure or absence of risk factors. Sometimes, the category with the highest level of exposure may be selected if you want to show protection. Also note that if the category with the lowest level of exposure has few observations, it may not be a good comparison group to use.
 <p>Reading</p> <p>1 hour</p>	<p>Practice Exercise #2</p> <ul style="list-style-type: none"> ● Keep participants in the same groups or pairs as in the previous exercise. ● Ask them to spend approximately 1 hour completing the exercise. ● Make sure you are available to answer any questions during the exercise.
 <p>Group Discussion</p>  <p>Flip Chart</p> <p>30 Minutes</p>	<p>Debrief</p> <p>Review each group or pairs' answers. If time permits, ask participants to present their answers to the entire class by drawing table shells on a flipchart.</p> <p>Possible answers and tables include:</p> <p>Section 1 – Case Study</p> <ol style="list-style-type: none"> 1. Use the space below to create at least three table shells to prepare for univariable analyses:

Duration/Session Type	What to Do/What to Say	
	Demographic Characteristics: Sex	
	Sex	Percentage (%) (Standard Error [SE])
	Male	
	Female	
	Total	
	Demographic Characteristics: Education – High School or More	
	Race	Percentage (%) (Standard Error [SE])
	Non-Hispanic white	
	Non-Hispanic black	
	Hispanic	
	Other	
	Blood Pressure Categories²	
	BP Levels	Percentage (%) (Standard Error [SE])
	Normotensive	
	Prehypertensive	
	Hypertensive	
2. Use the space below to create at least two table shells to		

² Could be prevalence, treatment or control of hypertension. *Prevalence: Average blood pressure $\geq 140/90$ mmHg or reported current use of blood pressure-lowering medication. Treatment: An answer of "yes" to the question "Are you currently taking medication to lower your blood pressure?" Among those with hypertension (average systolic blood pressure ≥ 140 mmHg, average diastolic pressure ≥ 90 mmHg, or current medication use). Control: Average treated blood pressure $< 140/90$ mmHg on examination among all persons with hypertension.*

Duration/Session Type	What to Do/What to Say																																																																																																				
	<p>prepare for bivariable analyses.</p> <p>Outcome: Hypertension</p> <table border="1" data-bbox="492 394 1466 716"> <thead> <tr> <th data-bbox="492 394 706 443"><u>Exposure:</u></th> <th colspan="3" data-bbox="706 394 1073 443"></th> <th colspan="3" data-bbox="1073 394 1466 443"></th> </tr> <tr> <th data-bbox="492 443 706 506"><u>Sex</u></th> <th colspan="3" data-bbox="706 443 1073 506">Yes</th> <th colspan="3" data-bbox="1073 443 1466 506">No</th> </tr> <tr> <th data-bbox="492 506 706 562"></th> <th data-bbox="706 506 808 562">N*</th> <th data-bbox="808 506 915 562">%</th> <th data-bbox="915 506 1073 562">95% CI</th> <th data-bbox="1073 506 1180 562">N*</th> <th data-bbox="1180 506 1287 562">%</th> <th data-bbox="1287 506 1466 562">95% CI</th> </tr> </thead> <tbody> <tr><td data-bbox="492 562 706 619"></td><td></td><td></td><td></td><td></td><td></td><td></td></tr> <tr><td data-bbox="492 619 706 676"></td><td></td><td></td><td></td><td></td><td></td><td></td></tr> <tr><td data-bbox="492 676 706 716"></td><td></td><td></td><td></td><td></td><td></td><td></td></tr> </tbody> </table> <table border="1" data-bbox="492 758 1466 1108"> <thead> <tr> <th data-bbox="492 758 706 806"><u>Exposure:</u></th> <th colspan="3" data-bbox="706 758 1073 806"></th> <th colspan="3" data-bbox="1073 758 1466 806"></th> </tr> <tr> <th data-bbox="492 806 706 905"><u>Race/Ethnic Group</u></th> <th colspan="3" data-bbox="706 806 1073 905">Yes</th> <th colspan="3" data-bbox="1073 806 1466 905">No</th> </tr> <tr> <th data-bbox="492 905 706 961"></th> <th data-bbox="706 905 808 961">N*</th> <th data-bbox="808 905 915 961">%</th> <th data-bbox="915 905 1073 961">95% CI</th> <th data-bbox="1073 905 1180 961">N*</th> <th data-bbox="1180 905 1287 961">%</th> <th data-bbox="1287 905 1466 961">95% CI</th> </tr> </thead> <tbody> <tr><td data-bbox="492 961 706 1018"></td><td></td><td></td><td></td><td></td><td></td><td></td></tr> <tr><td data-bbox="492 1018 706 1075"></td><td></td><td></td><td></td><td></td><td></td><td></td></tr> <tr><td data-bbox="492 1075 706 1108"></td><td></td><td></td><td></td><td></td><td></td><td></td></tr> </tbody> </table> <p>Facilitator Note: The first assessment of a cross-sectional study/survey is descriptive.</p> <p>The prevalence of a disease is the proportion of diseased individuals in a population. (i.e., prevalence = (cases)/(total population))</p> <table border="1" data-bbox="492 1482 1466 1656"> <thead> <tr> <th data-bbox="492 1482 764 1524"></th> <th colspan="3" data-bbox="764 1482 1466 1524">Outcome</th> </tr> </thead> <tbody> <tr> <td data-bbox="492 1524 764 1566">Exposure</td> <td data-bbox="764 1524 1037 1566">50 (a)</td> <td data-bbox="1037 1524 1263 1566">400 (b)</td> <td data-bbox="1263 1524 1466 1566">450</td> </tr> <tr> <td data-bbox="492 1566 764 1608"></td> <td data-bbox="764 1566 1037 1608">50 (c)</td> <td data-bbox="1037 1566 1263 1608">500 (d)</td> <td data-bbox="1263 1566 1466 1608">550</td> </tr> <tr> <td data-bbox="492 1608 764 1656"></td> <td data-bbox="764 1608 1037 1656">100</td> <td data-bbox="1037 1608 1263 1656">900</td> <td data-bbox="1263 1608 1466 1656"></td> </tr> </tbody> </table> <p>$P_1 = a/a+b = 50/450 = 11\%$ prevalence of outcome among people who are not exposed. $P_0 = c/c+d = 50/550 = 9\%$ prevalence of outcome among people who are exposed.</p>	<u>Exposure:</u>							<u>Sex</u>	Yes			No				N*	%	95% CI	N*	%	95% CI																						<u>Exposure:</u>							<u>Race/Ethnic Group</u>	Yes			No				N*	%	95% CI	N*	%	95% CI																							Outcome			Exposure	50 (a)	400 (b)	450		50 (c)	500 (d)	550		100	900	
<u>Exposure:</u>																																																																																																					
<u>Sex</u>	Yes			No																																																																																																	
	N*	%	95% CI	N*	%	95% CI																																																																																															
<u>Exposure:</u>																																																																																																					
<u>Race/Ethnic Group</u>	Yes			No																																																																																																	
	N*	%	95% CI	N*	%	95% CI																																																																																															
	Outcome																																																																																																				
Exposure	50 (a)	400 (b)	450																																																																																																		
	50 (c)	500 (d)	550																																																																																																		
	100	900																																																																																																			

Duration/Session Type	What to Do/What to Say																																														
	<p>Section 2 – Case Study</p> <p>1. Use the following template to list the variable pairs for which you will test a statistical association.</p> <table border="1" data-bbox="492 453 1463 720"> <thead> <tr> <th data-bbox="492 453 824 499">Statistical Test</th> <th data-bbox="824 453 1463 499">Variables to Assess</th> </tr> </thead> <tbody> <tr> <td data-bbox="492 499 824 606">chi-square</td> <td data-bbox="824 499 1463 606"><i>Race/Ethnicity vs. Hypertension prevalence</i></td> </tr> <tr> <td data-bbox="492 606 824 720">t-test</td> <td data-bbox="824 606 1463 720"><i>Gender by Continuous blood pressure</i></td> </tr> </tbody> </table> <p>2. Fill out the following tables to prepare for calculating measures of association.</p> <p>Outcome Variable: Hypertension</p> <table border="1" data-bbox="492 919 1463 1224"> <thead> <tr> <th data-bbox="492 919 732 1052">Exposure Variable:</th> <th data-bbox="732 919 966 1052"></th> <th data-bbox="966 919 1192 1052"></th> <th data-bbox="1192 919 1463 1052"></th> </tr> <tr> <td data-bbox="492 1052 732 1098">Sex</td> <td data-bbox="732 1052 966 1098">Yes</td> <td data-bbox="966 1052 1192 1098">No</td> <td data-bbox="1192 1052 1463 1098"></td> </tr> </thead> <tbody> <tr> <td data-bbox="492 1098 732 1144">Male</td> <td data-bbox="732 1098 966 1144"></td> <td data-bbox="966 1098 1192 1144"></td> <td data-bbox="1192 1098 1463 1144"></td> </tr> <tr> <td data-bbox="492 1144 732 1190">Female</td> <td data-bbox="732 1144 966 1190"></td> <td data-bbox="966 1144 1192 1190"></td> <td data-bbox="1192 1144 1463 1190"></td> </tr> <tr> <td data-bbox="492 1190 732 1224"></td> <td data-bbox="732 1190 966 1224"></td> <td data-bbox="966 1190 1192 1224"></td> <td data-bbox="1192 1190 1463 1224"></td> </tr> </tbody> </table> <p>PR = POR = $\chi^2 =$, $df =$, $p =$</p> <p>Outcome Variable: Hypertension</p> <table border="1" data-bbox="492 1455 1463 1797"> <thead> <tr> <th data-bbox="492 1455 771 1587">Exposure Variable:</th> <th data-bbox="771 1455 979 1587"></th> <th data-bbox="979 1455 1192 1587"></th> <th data-bbox="1192 1455 1463 1587"></th> </tr> <tr> <td data-bbox="492 1587 771 1633"><i>Race</i></td> <td data-bbox="771 1587 979 1633">Yes</td> <td data-bbox="979 1587 1192 1633">No</td> <td data-bbox="1192 1587 1463 1633"></td> </tr> </thead> <tbody> <tr> <td data-bbox="492 1633 771 1713">Non-Hispanic white</td> <td data-bbox="771 1633 979 1713"></td> <td data-bbox="979 1633 1192 1713"></td> <td data-bbox="1192 1633 1463 1713"></td> </tr> <tr> <td data-bbox="492 1713 771 1759">Other</td> <td data-bbox="771 1713 979 1759"></td> <td data-bbox="979 1713 1192 1759"></td> <td data-bbox="1192 1713 1463 1759"></td> </tr> <tr> <td data-bbox="492 1759 771 1797"></td> <td data-bbox="771 1759 979 1797"></td> <td data-bbox="979 1759 1192 1797"></td> <td data-bbox="1192 1759 1463 1797"></td> </tr> </tbody> </table> <p>PR = POR =</p>	Statistical Test	Variables to Assess	chi-square	<i>Race/Ethnicity vs. Hypertension prevalence</i>	t-test	<i>Gender by Continuous blood pressure</i>	Exposure Variable:				Sex	Yes	No		Male				Female								Exposure Variable:				<i>Race</i>	Yes	No		Non-Hispanic white				Other							
Statistical Test	Variables to Assess																																														
chi-square	<i>Race/Ethnicity vs. Hypertension prevalence</i>																																														
t-test	<i>Gender by Continuous blood pressure</i>																																														
Exposure Variable:																																															
Sex	Yes	No																																													
Male																																															
Female																																															
Exposure Variable:																																															
<i>Race</i>	Yes	No																																													
Non-Hispanic white																																															
Other																																															

Duration/Session Type	What to Do/What to Say																																		
	<p>$\chi^2 =$, $df=$, $p=$</p> <p>Outcome Variable: Hypertension</p> <table border="1" data-bbox="492 428 1468 730"> <thead> <tr> <th data-bbox="492 428 760 558">Exposure Variable:</th> <th data-bbox="760 428 954 558"></th> <th data-bbox="954 428 1170 558"></th> <th data-bbox="1170 428 1468 558"></th> </tr> </thead> <tbody> <tr> <td data-bbox="492 558 760 600"><i>Education</i></td> <td data-bbox="760 558 954 600">Yes</td> <td data-bbox="954 558 1170 600">No</td> <td data-bbox="1170 558 1468 600"></td> </tr> <tr> <td data-bbox="492 600 760 642"><12 Grade</td> <td data-bbox="760 600 954 642"></td> <td data-bbox="954 600 1170 642"></td> <td data-bbox="1170 600 1468 642"></td> </tr> <tr> <td data-bbox="492 642 760 684">>12 Grade</td> <td data-bbox="760 642 954 684"></td> <td data-bbox="954 642 1170 684"></td> <td data-bbox="1170 642 1468 684"></td> </tr> </tbody> </table> <p>PR = POR = $\chi^2 =$, $df=$, $p=$</p> <p>Facilitator Note: The second type of analysis in cross-sectional studies/surveys is analytic (after descriptive).</p> <p>Several tests can be used to test association between variables. Chi square tests can be used to assess general significance, but the test generally does not provide information on the strength of the association. Chi square tests are used with nominal/categorical variables and the cell counts should be greater than 5.</p> <p>Prevalence ratio and prevalence odds ratio can be used to test for associations and provide an assessment on the magnitude of the association.</p> <p>Example:</p> <table border="1" data-bbox="492 1560 1468 1734"> <thead> <tr> <th data-bbox="492 1560 764 1602"></th> <th colspan="3" data-bbox="764 1560 1468 1602">Outcome</th> </tr> </thead> <tbody> <tr> <td data-bbox="492 1602 764 1644">Exposure</td> <td data-bbox="764 1602 1037 1644">50 (a)</td> <td data-bbox="1037 1602 1289 1644">400 (b)</td> <td data-bbox="1289 1602 1468 1644">450</td> </tr> <tr> <td data-bbox="492 1644 764 1686"></td> <td data-bbox="764 1644 1037 1686">50 (c)</td> <td data-bbox="1037 1644 1289 1686">500 (d)</td> <td data-bbox="1289 1644 1468 1686">550</td> </tr> <tr> <td data-bbox="492 1686 764 1728"></td> <td data-bbox="764 1686 1037 1728">100</td> <td data-bbox="1037 1686 1289 1728">900</td> <td data-bbox="1289 1686 1468 1728"></td> </tr> </tbody> </table> <p>$P1= a/a+b= 50/450 = 11\%$ prevalence of outcome among people who are not exposed. $P0= c/c+d = 50/550 = 9\%$ prevalence of outcome</p>			Exposure Variable:				<i>Education</i>	Yes	No		<12 Grade				>12 Grade					Outcome			Exposure	50 (a)	400 (b)	450		50 (c)	500 (d)	550		100	900	
Exposure Variable:																																			
<i>Education</i>	Yes	No																																	
<12 Grade																																			
>12 Grade																																			
	Outcome																																		
Exposure	50 (a)	400 (b)	450																																
	50 (c)	500 (d)	550																																
	100	900																																	

Duration/Session Type	What to Do/What to Say																																																																				
	<p>among people who are exposed.</p> <p>$POR = ad/bc = (50 \times 500) / (400 \times 50) = 1.25.$ $PR = (a/N_1) / (c/N_0) = (50/450) / (50/550) = 1.2.$ The proportion of people with the outcome is 1.25 or 1.2 times greater if the person is unexposed compared to those exposed.</p> <p>Section 3 – Case Study:</p> <ol style="list-style-type: none"> List at least one potential confounder for the hypertension study. <i>Age, sex, education, race/ethnicity</i> What table shells do you need to prepare to assess for confounding and effect modification? Use the space below to create at least two table shells. <i>Is gender a risk factor for hypertension? (Among nonobese)</i> <p style="text-align: center;">Hypertension</p> <table border="1" data-bbox="492 1003 1463 1318"> <thead> <tr> <th rowspan="2">Gender</th> <th colspan="3">Yes</th> <th colspan="3">No</th> </tr> <tr> <th>N*</th> <th>%</th> <th>95% CI</th> <th>N*</th> <th>%</th> <th>95% CI</th> </tr> </thead> <tbody> <tr> <td>Male</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>Female</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> </tbody> </table> <p>*Unweighted N PR = POR =</p> <p style="text-align: center;">Hypertension</p> <table border="1" data-bbox="492 1493 1463 1843"> <thead> <tr> <th rowspan="2">Weight Classification</th> <th colspan="3">Yes</th> <th colspan="3">No</th> </tr> <tr> <th>N*</th> <th>%</th> <th>95% CI</th> <th>N*</th> <th>%</th> <th>95% CI</th> </tr> </thead> <tbody> <tr> <td>Obese</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>Nonobese</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> </tbody> </table>	Gender	Yes			No			N*	%	95% CI	N*	%	95% CI	Male							Female														Weight Classification	Yes			No			N*	%	95% CI	N*	%	95% CI	Obese							Nonobese													
Gender	Yes			No																																																																	
	N*	%	95% CI	N*	%	95% CI																																																															
Male																																																																					
Female																																																																					
Weight Classification	Yes			No																																																																	
	N*	%	95% CI	N*	%	95% CI																																																															
Obese																																																																					
Nonobese																																																																					

Duration/Session Type	What to Do/What to Say																																																																				
	<p>*Unweighted N PR = POR =</p> <p><i>Does gender modify the association between obesity and hypertension?</i></p> <p style="text-align: center;">Hypertension by Weight Classification - Gender: Males</p> <table border="1" data-bbox="492 726 1468 1077"> <thead> <tr> <th rowspan="2">Weight Classification</th> <th colspan="3">Yes</th> <th colspan="3">No</th> </tr> <tr> <th>N*</th> <th>%</th> <th>95% CI</th> <th>N*</th> <th>%</th> <th>95% CI</th> </tr> </thead> <tbody> <tr> <td>Obese</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>Nonobese</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> </tbody> </table> <p>*Unweighted N PR = POR =</p> <p style="text-align: center;">Hypertension by Weight Classification - Gender: Females</p> <table border="1" data-bbox="492 1350 1468 1701"> <thead> <tr> <th rowspan="2">Weight Classification</th> <th colspan="3">Yes</th> <th colspan="3">No</th> </tr> <tr> <th>N*</th> <th>%</th> <th>95% CI</th> <th>N*</th> <th>%</th> <th>95% CI</th> </tr> </thead> <tbody> <tr> <td>Obese</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>Nonobese</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> </tbody> </table> <p>*Unweighted N PR = POR =</p>	Weight Classification	Yes			No			N*	%	95% CI	N*	%	95% CI	Obese							Nonobese														Weight Classification	Yes			No			N*	%	95% CI	N*	%	95% CI	Obese							Nonobese													
Weight Classification	Yes			No																																																																	
	N*	%	95% CI	N*	%	95% CI																																																															
Obese																																																																					
Nonobese																																																																					
Weight Classification	Yes			No																																																																	
	N*	%	95% CI	N*	%	95% CI																																																															
Obese																																																																					
Nonobese																																																																					



Duration/Session Type	What to Do/What to Say
	<p>You may ask participants an additional question: Is gender in the pathway between obesity and hypertension? <i>No</i></p> <p>Note: When assessing confounding and effect modification, consider stratification of variables (e.g., age group, gender, etc.) to assess the primary relationship (i.e., obesity and hypertension). Stratification allows you to observe relationships beyond the crude association.</p>


SKILL ASSESSMENT

Total Estimated Time: 2 hours

Skill Assessment: 1 ½ hours


Debrief: 30 minutes

Duration/Session Type	What to Do/What to Say
 <p>Activity</p>  <p>Prepare</p> <p>1 ½ Hours</p>	<p>Skill Assessment</p> <ul style="list-style-type: none"> • The skill assessment contains four sections which is a compilation of all skills learned in the module. Participants will use the information you provide them on an NCD study to complete the skill assessment for this module and the other two data analysis modules. • Participants are encouraged to complete the skill assessment individually; however, if someone does not have an NCD study to work with, you can pair him/her up with another colleague. • Distribute the background information participants will need to complete the skill assessment for a study in their own country, such as: <ul style="list-style-type: none"> ○ Sample questionnaires ○ A description of the study including the information necessary to complete all sections, for example, the study question that needs to be answered. • Ask participants to take out their Activity Workbook and briefly explain the assignment. • Observe participants completing the assignment and be available to answer any questions.

Duration/Session Type	What to Do/What to Say
 <p>Group Discussion 30 Minutes</p>	<p>Debrief You may review participants' responses after each section or at the end of the skill assessment.</p>

CONCLUSION

Total Estimated Time: 15 minutes

Duration/Session Type	What to Do/What to Say
 <p>Group Discussion 15 Minutes</p>	<p>Conclusion</p> <ul style="list-style-type: none"> • Ask participants to provide some of the main points they learned in the module. • Ask participants for their reactions to what they learned in the training and how they will apply the skills when they return to their job. Ask what they will do <u>differently</u> after applying the information in this module.

Appendix A

Sample Agenda for Data Analysis Workshop

Day 1: Introduction to Workshop
Creating Data Analysis Plan module

Estimated Time	Activity	Facilitator(s)
8:30 – 9:00	Introduce participants and facilitators, and provide overview of workshop	
9:00 – 9:35	Participants read: Section 1: Introduction, Section 2: Overview of Data Analysis, and first half of Section 3: Analysis Plan	
9:35 – 9:45	Facilitator leads brief discussion on readings	
9:45 – 10:45	Practice Exercise #1 , including review of answers	
10:45 – 11:00	TEA/COFFEE BREAK	
11:00 – 11:40	Participants read how to prepare table shells	
11:40 – 12:00	Facilitator leads brief discussion on readings	
12:00 – 13:00	Practice Exercise #2	
13:00 – 14:00	LUNCH	
14:00 – 14:30	Facilitator leads review of Practice Exercise #2 answers	
14:30 – 16:30	Skill Assessment , including review of answers (Participants take their own 15-minute break)	
16:30 – 16:45	Facilitator concludes module	

Day 2: Managing Data module

Estimated Time	Activity	Facilitator(s)
8:30 – 8:40	Distribute Managing Data module Facilitator leads brief introduction to <i>Managing Data</i> module	
8:40 – 8:55	Participants read: Section 1: Introduction and Section 2: Overview of Data Management	
8:55 – 9:05	Participants read: Section 3: Data Dictionary	
9:05 – 9:20	Facilitator leads brief discussion on readings	
9:20 – 10:00	Practice Exercise #1 , including review of responses	
10:00 – 10:15	BREAK	
10:15 – 11:15	Skill Assessment #1 , including review of response	
11:15 – 11:35	Participants read: Section 4: Overview to cleaning data; common errors, duplicate records	
11:35 – 11:50	Facilitator leads brief discussion on readings	
11:50 – 12:30	Practice Exercise #2 , including review of responses	
12:30 – 13:00	Participants read: Missing data, miscodes, out-of-range data	
13:00 – 14:00	LUNCH	
14:00 – 14:30	Facilitator leads brief discussion on readings	
14:30 – 15:30	Practice Exercise #3 , including review of responses (Participants take their own 15-minute break)	
15:30 – 16:30	Skill Assessment #2 , including review of response	
16:30 – 16:45	Facilitator concludes module	

Day 3: Analyzing and Interpreting Large Datasets module

Estimated Time	Activity	Facilitator(s)
8:30 – 8:45	Facilitator lead brief review of previous days' activities	
8:45 – 8:50	Facilitator introduces <i>Analyzing Large Data Sets</i> module	
8:50 – 9:15	Participants read Sections 1 and 2	
9:15 – 9:25	Facilitator leads overview of readings	
9:25 – 9:55	Participants read part of Section 3 on univariable analysis and complete two brief activities	
9:55 – 10:15	BREAK	
10:15 – 10:30	Facilitator leads discussion on readings and reviews statistical software commands	
10:30 – 11:40	Practice Exercise #1 , including review of responses/answers	
11:40 – 12:00	Participants read bivariable analysis and complete 5 brief activities	
12:00 – 12:15	Facilitator leads discussion on readings and reviews statistical software commands	
12:15 – 13:05	Practice Exercise #2 , including review of responses/answers	
13:05 – 14:05	LUNCH	
14:05 – 14:40	Participants read part of Section 4: Overview to analytic statistics, concepts of association, statistical significance testing and confidence intervals	
14:40 – 14:55	Facilitator leads overview of readings and reviews statistical software commands	
14:55 – 16:15	Practice Exercise #3 , including review of responses/answers	

Day 4: Analyzing and Interpreting Large Datasets module

Estimated Time	Activity	Facilitator(s)
8:30 – 9:05	Participants read about stratified analysis, EFFECT MODIFICATION, and confounding	
9:05 – 9:20	Facilitator leads discussion on readings from previous day	
9:20 – 10:25	Practice Exercise #4 and review of answers	
10:25 – 10:40	COFFEE/TEA BREAK	
10:40 – 11:00	Participants read about interpreting and reporting findings	
11:00 – 11:15	Facilitator leads discussion on readings and reviews statistical software commands	
11:15 – 12:15	Practice Exercise #5 and review of answers	
12:15 – 13:00	Begin Skills Assessment	
13:00 – 14:00	LUNCH	
14:00 – 16:30	Skill Assessment	

Day 5: Analyzing and Interpreting Large Datasets module

Estimated Time	Activity	Facilitator(s)
8:30 – 10:15	Complete Skill Assessment and present findings	
10:15 – 10:30	COFFEE/TEA BREAK	
10:30 – 12:00	Complete presentations; Facilitator leads conclusion of the module/course	

Appendix B

Sample Review Questions for Sections 1 and 2 and First Half of Section 3:

1. What are some benefits of creating an analysis plan?

Possible answer:

An analysis plan helps you think through the data you will collect, what you will use it for, and how you will analyze it. Creating an analysis plan is an important way to ensure you collect all the data you need and that you use all the data you collect.

2. An FETP resident created an analysis plan by including the following:
 - Research question(s) and/or hypotheses
 - Dataset to be used
 - Table shells for univariable analysis

What additional components of the analysis plan should she have included?

Answer:

- *Inclusion/exclusion criteria*
- *Variable to be used in the main analysis*
- *Statistical methods and software to be used*
- *Table shells for bivariable and stratified analysis*

Review the answers to the types of statistical data activity by asking questions such as:

1. What is an example of nominal data?
2. What is an example of ordinal data?
3. What is an example of interval data?
4. What is an example of ratio data?

Sample Review for Section 4:

Ask for volunteers to provide examples of:

- A univariable table
- A bivariable table