

# ACTIVITY WORKBOOK



## Managing Data

Created: 2013





*Managing Data.* Atlanta, GA: Centers for Disease Control and Prevention (CDC), 2013.

### Activity

#### Skill Assessment #1

Estimated Time: 40 minutes

#### **Background:**

You will work in pairs to create part of a data dictionary, based on information provided by your facilitator.

#### **Instructions:**

1. Your facilitator or mentor will provide you with a partially completed data dictionary for an NCD study conducted in your country.
2. Work with your colleague to complete the variables for the data dictionary. You can use the dictionary handed to you by your facilitator, your computer, or the space below. At the end of the exercise, you will share your work with the rest of the class.

**Let your facilitator know when you have completed the Skill Assessment.**

## Activity

## Skill Assessment #2

Estimated Time: 45 minutes

**Background:**

You will work individually to clean data from your country's dataset. (If you did not bring "dirty" data to class, you can pair up with a colleague to complete the assessment.)

**Instructions:**

1. Your instructor will provide you with a dirty dataset for which you will clean and the supporting questionnaires, if available.
2. Your task is to identify errors in the dataset, including missing data, duplications, miscodes and outliers. Correct the data you can, and use the tables below to document your resolution.

For Duplicate Records:

Number	Primary ID	Secondary ID	Problem	Record IDs Affected and Resolution

For All Other Errors:

<b>Number</b>	<b>Variable name</b>	<b>Format</b>	<b>Problem</b>	<b>Record IDs affected and resolution</b>

**Let your facilitator know when you have completed the Skill Assessment.**