Department of Epidemiology
School of Public Health
University of California, Los Angeles

Problem Two (Data Analysis)

Complete Chapters 1&2 in the Software Training Manual
and study the Csurvey Manual before doing this problem

Please note that some of the material will not be covered until later in the course. Complete what you can, and then wait for the remainder to be addressed.

This is a graded homework. You may get assistance from your assigned team partner in doing the analysis, but the write-up and interpretation must represent your work alone.

Assume that you are consulting with the government of Niger, a country of 10.6 million people in West Africa. Specifically, the government officials are concerned with a rise in sexually-transmitted diseases among young adults in Niamey, the largest city in the country with a population of nearly 660,000 people (2008 Estimate: 659,055). Rather than focusing on men, the officials would like to know the levels of human immunodeficiency virus (HIV) infection and syphilis among women, aged 20-29 years. An earlier national survey done in 2002 [1] found that 3.4 percent of Niger women aged 20-29 in urban areas were HIV positive. Because it was a national survey of the total population, however, the findings were more intriguing than definitive. Other investigators have suggested that the prevalence of syphilis is about 9 percent, although based on mathematical models and clinic data. Thus the government is asking you to conduct a more detailed survey.

You decide to do a two-stage cluster survey of 1,200 sexually-active women, aged 20-29 years, living in Niamey, Niger. Approximately 93 percent of the women are married or in long-term relationships. The city is divided into three communes, with each further divided into districts: commune A has 78 districts, commune B has 95 districts, and commune C has 70 districts. Hence there are 213 districts in the city, with population data available for each. You decide to sample 30 clusters with probability proportional to size (PPS). In each cluster you decide to use the spin dial method to select two times a random start household. For each of the two times, you then select 20 eligible women assembled from the next nearest households. Thus for each of the 30 clusters, in two groups you will have selected 40 eligible women, aged 20-29, who are available on the day of the home-visit. You decide to collect saliva specimens for HIV assessment (S-HIV), using the method done in Thailand [2], and use the Rapid Syphilis Test (RST), field tested in Gambia [3], to determine the presence of syphilis. In addition to their age, you decide to ask the women about their age at first intercourse, interaction with casual (i.e., not married or living together) sexual partners, and condom use. Based on a demographic profile of the Niamey population, you estimate that there are 50,070 women, aged 20-29 in the city, 95 percent of whom are sexually active. Thus you plan to sample 2.5 percent of them. The last census for Niamey found that there were 3.6 persons per household. For planning purposes, a reasonable assumption is that there is no more than one sexually active 20-29 year old woman in a household.

The interview schedule is as shown at the end of this section. Recognizing the need to keep the interview short, each participant will be asked only five questions. In addition, each individual is asked to provide a saliva specimen and a small amount of blood via a finger-stick. An identification number is created by combining the cluster number (ranging from 1 to 30), and the

person number in the cluster (ranging from 1 to 40). At the end of the study, the saliva and blood specimens are taken to a laboratory for analysis. The saliva specimens are analyzed with one enzyme-linked immunosorbent assay (EIA) and then if positive, confirmed with a second EIA, different from the first, adhering to diagnostic Strategy II of the World Health Organization. The laboratory reports the findings as either HIV positive (both tests are reactive) or HIV negative (at least one test is non-reactive). The authors of the Thailand HIV saliva testing reported that under field conditions the test had a sensitivity and specificity of 98.0 and 99.4 percent, respectively. For the sake of this problem, there are no inconclusive laboratory findings. The blood is analyzed using an RST strip with recombinant *Treponema pallidum* antigens, requiring 100 µl of serum, separated from the blood cells. The RST strip is read as either positive or negative; again, for the sake of this problem, there are no inconclusive findings. Finally similar to the Gambia study [3], you assume that the sensitive and specificity of the RST test under field conditions is 75.0 and 95.2 percent, respectively.

The information is collected in an anonymous manner using headphones and recorders, with no identifying variables to link the results to any specific person other than an identification number. Each woman is given her identification number, however, so that she can find out her HIV and syphilis status. In about one to two weeks, the study personnel will send the results by ID number to a local testing and counseling center. The women can come to the center, state their ID number, and get the findings free of charge along with a follow-up diagnostic test. Counseling and notification of support services are also available for family members of the HIV and syphilis infected women who participated in the survey. You estimate, based on the study management forms, that about 12 percent of the women were not home at the time of the interview, and thus were replaced in the sample.

<table>
<tr><td colspan="3">Department of Epidemiology
University of California at Los Angeles (UCLA)
Los Angeles, CA USA

Schedule A Urban STD Survey of Women</td></tr>
<tr><td>1. ID No. __ __ __ __</td><td>Cluster No. __ __ and Person No. in Cluster __ __</td><td>2. Cluster No. __ __</td></tr>
<tr><td colspan="2">3. Current age __ __ years</td><td>4. Age at first sexual intercourse: [1] < 15 years [0] ≥ 15 years</td></tr>
<tr><td colspan="3">5. Recent sexual partner a casual partner? [1] Yes [0] No</td></tr>
<tr><td colspan="3">6. Consistent condom use with most recent partner? [1] Yes [0] No</td></tr>
<tr><td colspan="2">7. HIV (saliva): [1] positive [0] negative</td><td>8. Rapid Syphilis Test: [1] positive [0] negative</td></tr>
</table>

Analysis Team

The members of the class have been divided into 13 teams of two (and one of three) members (see the following table). The teams were mixed by medical background and academic experience in epidemiology. My intent is to have you work together only on some aspect of the analysis, but separately on the rest, including other aspects of the analysis, creation of forms and the report. By working together as a team, I assume that one will help teach the other so that both understand methodological details of the more complex analyses and how to use the software. By working separately on the individual part, the team members gain individual experience and credit for their achievements. If you need help with the problem, please consult your team member (only for the team portion) or Professor Frerichs.

Team	Student Name (1)	Student Name (2)	Student Name (3)
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			
11			
12			
13			

The Research Questions

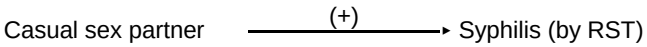
Your suspicion, based on the work of other investigators, is that the prevalence of syphilis and HIV infection are both associated with age at first intercourse, having casual sex partners and consistent condom use. You decide in your survey not to ask about marital status since the vast majority (i.e., est. 93%) are married or in long-term relationships, but rather to focus on the stated potential etiologic dimensions, assuming the women will respond in a factual manner. Some studies have shown that persons with syphilis are more likely to become infected with HIV, while others suggest that HIV infected persons are more likely to become infected with syphilis. With prevalence data you will not be able to determine which disease came first, but will be able to assess if there is an association. Your **broad objective** is to describe the occurrence of selected sexual intercourse-related variables, and diseases syphilis and HIV among sexually-active women, aged 20-29 years, now living in Niamey, Niger. Besides describing the findings, the **specific research questions** that you will want to answer are presented in the coming pages.

As you are considering the work, notice the number of questions that can be addressed in the analysis with only six subject-matter variables (i.e., variables 3 through 8). While the interview schedules for typical rapid surveys are limited to two pages, this still provides you with about 2-3 times the number of variables featured in *Problem Two*. As you will soon learn, limiting variables to those which can be listed on two pages is a blessing, not a curse.

The specific steps for completing *Problem Two* follow these research questions.

1. Are sexually-active women with a recent casual (i.e., not married or committed relationship) sex partner more likely to be infected with syphilis than women who did not have a recent casual partner?

***Implied Model*:**

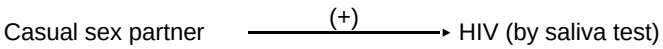


Explanation of Model

Having a casual sex partner is positively associated with the prevalence of syphilis

2. Are sexually-active women with a recent casual (i.e., not married or committed relationship) sex partner more likely to be infected with HIV than women who did not have a recent casual partner?

***Implied Model*:**

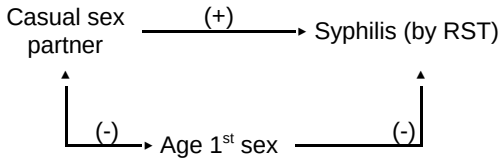


Explanation of Model

Having a casual sex partner is positively associated with the prevalence of HIV

3. Is the apparent relationship between having a casual sexual partner and syphilis explained by the potential confounding effects of age at first sexual intercourse?

***Implied Model (if "age 1st sex" is a confounding variable)* :**

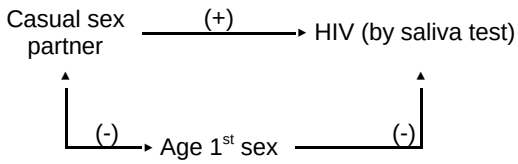


***Explanation of Model (this describes your hypothesis, and not necessarily your data)**:*

To be a confounding variable, *age at first sex* would have to have three characteristics. It is: 1) an independent risk factor of the outcome variable (i.e., syphilis), 2) associated with the exposure variable (i.e., having a casual sex partner), and 3) is not a consequence of the exposure variable (i.e., having a recent casual sex partner does not make a person have early sexual intercourse). The associations between *Age at first sex* and the two other variables are negative, implying having a casual sex partner is associated with younger age at first sex (i.e., higher "casual sex", lower "age first sex"), and younger age at first sex is associated with increased syphilis (i.e., lower "age first sex", higher "syphilis").

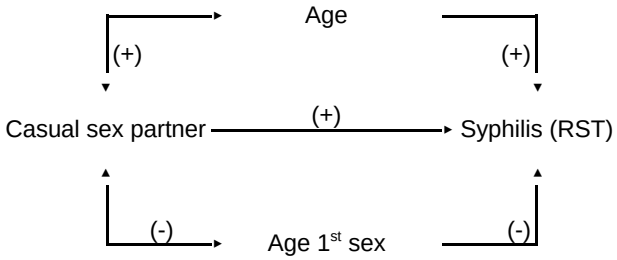
4. Is the apparent relationship between having a casual sexual partner and HIV explained by the potential confounding effects of age at first sexual intercourse?

Implied Model (if "age 1st sex" is a confounding variable) :



5. After controlling for the potential confounding effect of current age (i.e., ranging from 20-29 years) and age at first sexual intercourse, are sexually-active women with a recent casual sex partner more likely to be infected with syphilis than women who did not have a recent casual partner?

***Implied Model (if "age" and "age 1st sex" are confounding variables)*:**

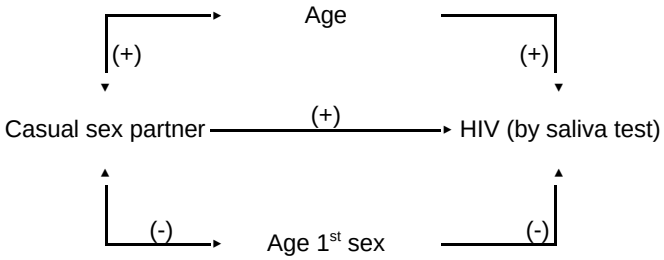


Explanation of Model

The implication of the model is that *age* and *age first sex* are associated with, but not a consequence of *casual sex partner* and that *age* and *age first sex* are both independently associated with the prevalence of syphilis. If these assumptions do not hold, then the variables should not be included in the analysis as confounding variables.

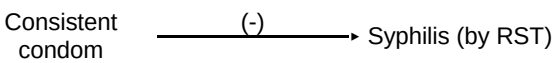
6. After controlling for the potential confounding effect of current age (i.e., ranging from 20-29 years) and age at first sexual intercourse, are sexually-active women with a recent casual sex partner more likely to be infected with HIV than women who did not have a recent casual partner?

***Implied Model (if "age" and "age 1st sex" are confounding variables)*:**



7. Is consistently using a condom with the most recent sexual partner protective against syphilis infection?

***Implied Model*:**

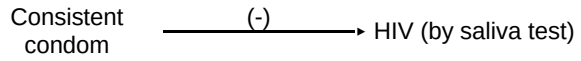


Explanation of Model

Consistent condom users are less likely to be infected with syphilis.

8. Is consistently using a condom with the most recent sexual partner protective against HIV infection?

***Implied Model*:**

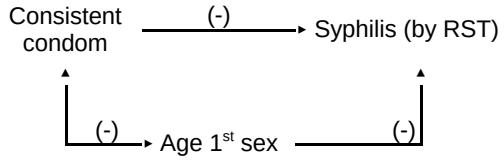


Explanation of Model

Consistent condom users are less likely to be infected with HIV.

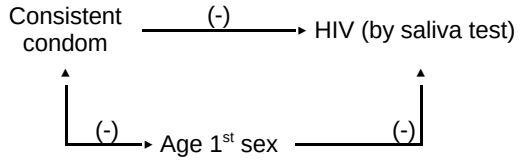
9. Is the apparent relationship between consistent condom use and existing syphilis explained by the potential confounding effects of age at first sexual intercourse?

***Implied Model (if "age 1st sex" is a confounding variable)* :**



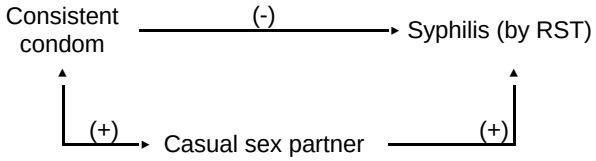
10. Is the apparent relationship between consistent condom use and existing HIV infection explained by the potential confounding effects of age at first sexual intercourse?

***Implied Model (if "age 1st sex" is a confounding variable)* :**



11. Is the apparent relationship between consistent condom use and existing syphilis explained by the potential confounding or intermediate effects of recent sexual partner being a casual partner?

***Implied Model (if "age 1st sex" is a confounding variable)* :**



***Explanation of Model (this describes your hypothesis, and not necessarily your data)*:**

Casual sex partner may be a confounding variable or intermediate variable. While a confounding variable has three characteristics, an intermediate variable has only two. It is: 1) an independent risk factor of the outcome variable (i.e., existence of syphilis), and 2) is a consequence of the exposure variable (i.e., consistent condom use results in casual sex partner). If an intermediate variable, the model would suggest that some of the effect of consistent condom use on syphilis infection is direct and some is indirect via casual sex partner.

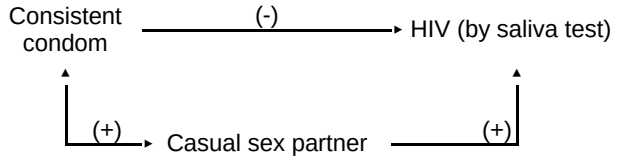
Policy Implications of an intermediate variable:

If consistent condom use is protective of syphilis infection, policies would need to be established to encourage

condom use. The analysis should indicate the anticipated benefits of consistently using condoms. To compare the relative contribution of the direct and indirect pathway from consistent condom use to the prevalence of syphilis, you would compare the effect of consistent condom use on syphilis by strata of "recent sexual partner a casual partner," thereby controlling for *casual sex partner*. Such control is typically done in etiological studies. Yet for policy purposes if *casual sex partner* is an intermediate variable, you should not control for it since this artificially reduces the true association between *consistent condom* and the existence of syphilis. Conversely if *casual sex partner* is a confounding variable, it would need to be controlled since confounders distort the truthful relationship between an exposure variable and the disease variable. For policy purposes, different from etiologic analyses, confounding variables are always controlled while intermediate variables are not. This is why you must determine what types of variables you are working with. To do so, keep in mind the three criteria of a confounding variable and the two criteria of an intermediate variable.

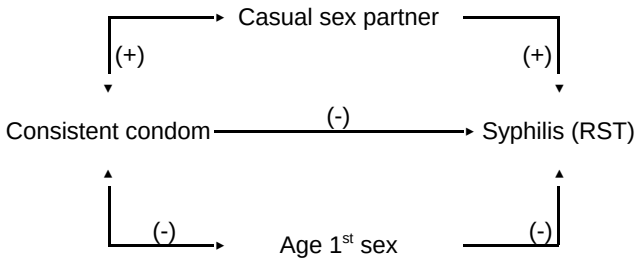
12. Is the apparent relationship between consistent condom use and the prevalence of HIV infection explained by the potential confounding or intermediate effects of recent sexual partner being a casual partner?

Implied Model (if "age 1st sex" is a confounding variable) **:**



13. After controlling for the potential effects of recent sex partner being a casual partner and age at first sexual intercourse, are sexually-active women who reported consistent condom use more likely to be infected with syphilis than women who did not report that their most recent partner consistently used a condom?

Implied Model (if "casual sex partner" and "age 1st sex" are confounding variables)**:**

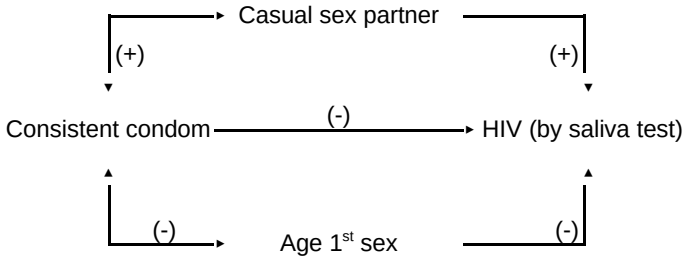


Explanation of Model

The implication of the model is that *casual sex partner* and *age first sex* are associated with, but not a consequence of *consistent condom* (i.e., are confounding rather than intermediate variables) and that *casual sex partner* and *age first sex* are both independently associated with the prevalence of syphilis. If these assumptions do not hold, then the variables should be reconsidered for the analysis.

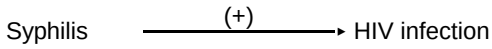
14. After controlling for the potential effects of recent sex partner being a casual partner and age at first sexual intercourse, are sexually-active women who reported consistent condom use more likely to be infected with HIV than women who did not report that their most recent partner consistently used a condom?

***Implied Model (if "casual sex partner" and "age 1st sex" are confounding variables)*:**



15. Are sexually active women with syphilis more likely to be infected with HIV than women with no syphilis?

***Implied Model*:**

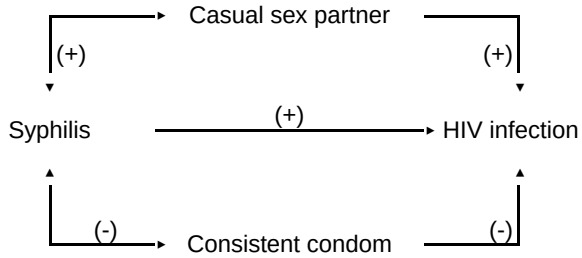


Explanation of Model

Having syphilis is positively associated with HIV infection

16. After controlling for the potential confounding effect of *casual sex partner* and *consistent condom*, are sexually active women with syphilis more likely to have HIV infection ?

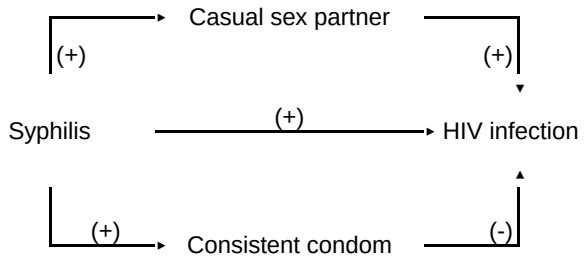
***Implied Model (if "casual sex partner" and "consistent condom" are confounding variables)*:**



Explanation of Model

Having the recent sexual partner be a casual partner is associated with syphilis infection, but is not a consequence of having syphilis. Having a casual sex partner is an independent risk factor for HIV infection. Consistently using a condom with the most recent sexual partner is negatively (i.e., preventive) associated with syphilis and is protective of HIV infection.

***Implied Model (if "casual sex partner" and "consistent condom" are intermediate variables)*:**



Explanation of Model

Women who know or sense they have syphilis are more likely to have sex with a causal partner and are more likely to consistently have their partners use a condom. Both "casual sex partner" and "consistent condom" are risk factors (i.e., one causal and other preventive) of HIV infection.

While typically you would adjust your data for the effects of a confounding variable, such an adjustment would usually not be done with an intermediate variable. Here you have possible confounding variables or intermediate variables, and most decide whether to adjust for neither, one or both. What should you do? The answer requires subject matter knowledge and/or assumptions as to the state (i.e., confounder or intermediate) of a variable. This is not a decision that can be made by a statistician as part of the analysis, but instead must be made by the knowledgeable investigator.

DATA

The **schedule** used to collect the data is shown on page 2. The short names for the variables are as follows:

Data Field Number	Variable Name
1	ID
2	cluster
3	age
4	sex15
5	casual
6	condom
7	SHIV
8	RST

For this problem, you will be using a **team-specific** data set, drawn randomly from a larger set representative of the population of women that might have been surveyed in Niamey, Niger. The data for 1,200 subjects done 13 times were created by Professor Frerichs as a teaching example (i.e., not appropriate to be published in a research journal) generated as *sex_prob1dataX* , where X is a number from 1 through 13, corresponding to the team number. The population data (all but 40 regions), necessary for selecting 30 clusters with probability proportionate to size, are in *NIGERHIV.csf* (i.e., a *Csurvey 2.0* data set). This file contains the names and population size of 173 of the 213 administrative regions in Niamey. Both files necessary for this problem are located at *http://www.ph.ucla.edu/epi/rapidsurveys/RScourse/RSproblems.html* on the Rapid Survey Course website in a single zip file, *RSCprob2teamX.zip* (where X varies from 1 through 13). Download only the appropriate zip file for your team (see table on page 3 for team number), save it in your

computer, and then unzip the file to gain access to *NIGERHIV.csf* and *sex_prob2dataX.mdb* (i.e., a *Epi Info* windows file).

Each student in the two-person team will use the same data set containing 1,200 sampled women. *Please work together with your team-mate on some of the tasks, but alone on the others.* It is unethical in this class to share your calculations and work on *Problem Two* with others outside your team. This is a graded undertaking and is intended to reflect solely on your performance (with some help on aspects of the analysis from your team-mate). Above all, it is intended to provide you with intellectual strength and insights for creating and conducting rapid surveys.

Before doing a survey you need to determine if the sample size being proposed is adequate for those planning to use the information. Next you need to create the interview schedule (already done) and management forms (to be done) to guide and assist the field staff. These forms, described in Chapter 2 of the *Software Training Manual* are:

Form 1 Map of Households

Form 2 Table of Random Numbers

Form 3 Household Disposition

Form 4 Study Disposition (two pages; Forms 4a and 4b)

You will need to create each of the forms for managing the survey. For *Problem Two*, you will also need to fill out Forms 1, 3, and 4 for two of the 30 clusters in your study (you select them).

To far, you have been given a general overview of what is to come. To complete *Problem* Two, you will need to address the following 26 instructions or questions and complete a **brief report** of your findings. The assignment will account for 35 percent of your course grade (50% is the final *Group Problem* and 15% was *Problem One*). *Problem Two* is due **Session 14** at the beginning of class. Please remember to work alone, although you may seek assistance of others on using the *Epi Info*, *CSurvey*, and *Stata* software.

SPECIFIC STEPS

INDIVIDUAL SECTION (Steps 1-6; this should reflect your individual creative thought; do not work with your team-mate(s) on the tasks in this section)

1. Create a written report with the items in *Specific Steps* section, including figures, and briefly describe the findings and policy implications. The sections of the report consist of...
    a. Introduction (prepare individually)
        i. Brief – do not include literature review, other than what was mentioned in the problem
    b. Methods (prepare individually)
    c. Results (prepare as a team and individually)
        i. Table 1 (analyze data individually and decide on table layout individually)
        ii. Table 2 (analyze data as a team, but decide on table layout individually)
        iii. Table 3 (analyze data as a team, but decide on table layout individually)
        iv. Figure 1 (analyze data as a team, but decide on figure layout individually)
        v. Figure 2 (analyze data as a team, but decide on figure layout individually)
    d. Discussion and conclusions (prepare individually)
    e. References (prepare individually)
        i. Include only those articles mentioned in the problem or obtained in this course (i.e.,

you might want to use Frerichs and Tar Tar, 1989 (**Reference 3**), and Frerichs, 1989 (**Reference 4**) to reference the rapid survey method).

f. Appendix (analyze data as a team, but decide on appendix layout individually)
   i. Table A
   ii. Table B
   iii. Table C
   iv. Form 1 Example
      (1) Form 1 Filled out for one cluster
      (2) Form 1 Filled out for a second cluster
   v. Form 2 Example
   vi. Form 3 Example
      (1) Form 3 Filled out for one cluster
      (2) Form 3 Filled out for a second cluster
   vii. Form 4 Example
      (1) Form 4 Filled out for two clusters
   viii. Table D
   ix. Table E

2. Create the four management forms for your survey.

   a. Use either the tables function in a word processing program such as *Word* (as described in Chapter 2 of the Software Training Manual) or in a spreadsheet program such as *Excel*.

   b. Form 2 can be created by using either *CSurvey 2.0* or a *Excel* spreadsheet program with random number generator (as described in Chapter 2 of the Software Training Manual). The maximum number of households in any of the 850 urban regions of Niamey, Niger between the center of the region and periphery is approximately 200. Use this information when creating Form 2.

3. Using *CSurvey 2.0* and the partial population file *NIGERHIV.csf*, enter the data for the remaining 40 administrative areas in Commune C (see table on page 18). Each of the two students in the team should individually enter the indicated information for his or her 20 administrative regions (one followed by the other – then share the completed NIGERHIV.csf file).

   a. Enter the appropriate information in the *Survey Parameter* section of *Csurvey 2.0*, based on the population information that was provided to you in the introduction of this problem.

      i. Hint: you will individually need to think your way through the process and do the calculations.
      ii. Re-save the *.csf file using the name NIGERHIV_1.csf for student 1 (this includes the individually-derived *Survey Parameter* information) and NIGERHIV_2.csf for student 2.
      iii. Print the *Survey Parameter* page and include it as **Table A** in the **Appendix** of your report.

   b. Select 30 cluster with probability proportionate to population size from the 213 urban

regions of Naimey, Niger as now tallied in *NIGERHIV.csf* for the 659,055 persons in the city.

c. Include your *CSurvey* lists of 213 administrative areas and of 30 selected clusters as **Tables B and C** in the **Appendix** of your report.

4. Fill out Forms 1, 3 and 4 for <u>two</u> selected clusters, keeping in mind the design within each cluster of selecting two random-start households.

   a. Use your imagination to describe what the field team likely found in each cluster (i.e., urban region of Niamey, Niger) when filling out (in pencil or ink) your created forms.
   b. Include these forms in the **Appendix** of your report.

5. Obtain from the zip file *sex_prob2dataX .mdb*, where X is a number from 1 through 13 corresponding to your team number), to be used with *Epi Info* (Windows) and, following conversion, use *sex_prob2dataX.dta* with *Stata*. Your one data set includes information on 1,200 eligible women, selected at two stages as 30 clusters and 40 women per cluster (selected in two groups).

6. Derive a frequency distribution of *age*, *sex15*, *casual*, *condom*, *SHIV* and *RST* and present in **Table 1** of your report.
   a. Include for all the 95% confidence interval
      i. Most are binomial variables (i.e., percentage)
      ii. *Age* in an equal interval variable (i.e., mean)
   b. Be sure to use proper descriptions of the variables, explained in a few words or short sentence, rather than the exact variable names.

TEAM SECTION (Steps 7-26; this section should be done in collaboration with your team member, so that both are using the same analyses for their individual work)

7. Determine using *CSurvey* if the sample size of an average of 40 eligible women in each of 30 clusters will be adequate to assess the prevalence of HIV infection in the study population within plus or minus 2 percentage points and the prevalence of syphilis infection, also within plus or minus 2 percentage points (see first section of *Problem Two* for details). Assume when using *CSurvey* that the value of the *design effect* (i.e., the variance of the cluster survey divided by the variance of a same-sized simple random sample) is "low" (i.e., around 2).
   a. Include *CSurvey* findings as **Table D** in the **Appendix** of your report
      i. Be sure to show the 90, 95 and 99 percent confidence limits for both the estimated HIV prevalence and syphilis prevalence
8. Using the assumed sensitivity and specificity of the saliva-based HIV test and the syphilis RST...
   a. derive the 95 percent confidence limits for the estimated <u>true</u> HIV prevalence and syphilis prevalence in study population (see first section of *Problem Two* for test characteristics).
      i. Present the confidence interval data for both estimated observe prevalence and estimated true prevalence for HIV and syphilis in **Table E** in the **Appendix** of the report

ii. The mathematical relationship between true prevalence and observed prevalence is covered in the *epidemiologic methods* three-course series. The material is summarized here.

$$P = \frac{m_1}{n} = \text{true prevalence}$$

$$P' = \frac{m}{n} = \text{observed prevalence}$$

$$SE = \frac{TP}{m_1} = \text{sensitivity of test}$$

$$SP = \frac{TN}{m_0} = \text{specificity of test}$$

		True Disease Status: D	True Disease Status: $\overline{D}$	
Test Result	+	TP	FP	m_1'
	−	FN	TN	m_0'
		m_1	m_0	n

$$\hat{P} = \frac{P' + SP - 1}{SE + SP - 1} = \text{estimated true prevalence}$$

9. Derive by two-year age groups, the point estimate and 95 percent confidence interval for the percent of the study population that is currently infected with HIV and syphilis.
   a. Present in **Figure 1** as a line graph.

10. Determine the point estimate and 95 percent confidence interval for the percent of syphilis and HIV infection among women who had their first sexual intercourse before age 15 years versus at 15 years and older.
    a. Present in **Figure 2** as a bar graph.

11. Answer the research question, "Are sexually-active women with a recent casual (i.e., not married or committed relationship) sex partner more likely to be infected with syphilis than women who did not have a recent casual partner?"
    a. Determine the crude prevalence ratio (i.e., derive as risk ratio but recognize the data are a measure of prevalence, not incidence) and crude odds ratio with both the *Complex Sample* module of *Epi Info* and with the survey modules of *Stata*.
       i. Include the 95% confidence interval
       ii. To what extent has cluster sampling increased the variances of the odds ratio and risk ratio above what they would have been if we had used simple random sampling?
    b. Present in **Table 2** for *Epi Info* and *Stata*, and in **Table 3** only for *Stata*.

12. Answer the research question, "Are sexually-active women with a recent casual sex partner more likely to be infected with HIV than women who did not have a recent casual partner?"
    a. Determine the crude prevalence ratio and crude odds ratio with both the *Complex Sample* module of *Epi Info* and with the survey modules of *Stata*.
       i. Include the 95% confidence interval
       ii. To what extent has cluster sampling increased the variances of the odds ratio and risk ratio above what they would have been if we had used simple random sampling?

b. Present in **Table 2** for *Epi Info* and *Stata*, and in **Table 3** only for *Stata*

13. Answer the research question, "Is the apparent relationship between having a casual sexual partner and syphilis explained by the potential confounding effects of age at first sexual intercourse?"
    a. View for yourself (but not to be included in the table) the stratum-specific odds and risk ratios
    b. Determine the adjusted odds ratio using the logistic function of *Stata* and adjusted risk ratio using the Poisson regression function of *Stata*.
        i. Include the 95% confidence intervals.
        ii. To what extent has cluster sampling increased the variances of the adjusted odds ratio and risk ratio above what they would have been if we had used simple random sampling?
    c. Present in **Table 3**

14. Answer the research question, "Is the apparent relationship between having a casual sexual partner and HIV explained by the potential confounding effects of age at first sexual intercourse?
    a. View for yourself (but not to be included in the table) the stratum-specific odds and risk ratios
    b. Determine the adjusted odds ratio using the logistic function of *Stata* and adjusted risk ratio using the Poisson regression function of *Stata*.
        i. Include the 95% confidence intervals.
        ii. To what extent has cluster sampling increased the variances of the adjusted odds ratio and risk ratio above what they would have been if we had used simple random sampling?
    c. Present in **Table 3**

15. Answer the research question, "After controlling for the potential confounding effect of current age (i.e., ranging from 20-29 years) and age at first sexual intercourse, are sexually-active women with a recent casual sex partner more likely to be infected with syphilis than women who did not have a recent casual partner?"
    a. View for yourself (but not to be included in the table) the stratum-specific odds and risk ratios
    b. Determine the adjusted odds ratio using the logistic function of *Stata* and adjusted risk ratio using the Poisson regression function of *Stata*.
        i. Include the 95% confidence intervals.
        ii. To what extent has cluster sampling increased the variances of the adjusted odds ratio and risk ratio above what they would have been if we had used simple random sampling?
    c. Present in **Table 3**

16. After controlling for the potential confounding effect of current age (i.e., ranging from 20-29 years) and age at first sexual intercourse, are sexually-active women with a recent casual sex partner more likely to be infected with HIV than women who did not have a recent casual partner?
    a. View for yourself (but not to be included in the table) the stratum-specific odds and risk ratios
    b. Determine the adjusted odds ratio using the logistic function of *Stata* and adjusted risk ratio using the Poisson regression function of *Stata*.
        i. Include the 95% confidence intervals.
        ii. To what extent has cluster sampling increased the variances of the adjusted odds ratio and risk ratio above what they would have been if we had used simple random sampling?
    c. Present in **Table 3**

17. Answer the research question, "Is consistently using a condom with the most recent sexual partner protective against syphilis infection?
    a. Determine the crude prevalence ratio and crude odds ratio with both the *Complex Sample* module of *Epi Info* and with the survey modules of *Stata*.
        i. Include the 95% confidence interval
        ii. To what extent has cluster sampling increased the variances of the odds ratio and risk ratio above what they would have been if we had used simple random sampling?
    b. Present in **Table 2** for *Epi Info* and *Stata*, and in **Table 3** only for *Stata*

18. Answer the research question, "Is consistently using a condom with the most recent sexual partner protective against HIV infection? "
    a. Determine the crude prevalence ratio and crude odds ratio with both the *Complex Sample* module of *Epi Info* and with the survey modules of *Stata*.
        i. Include the 95% confidence interval
        ii. To what extent has cluster sampling increased the variances of the odds ratio and risk ratio above what they would have been if we had used simple random sampling?
    b. Present in **Table 2** for *Epi Info* and *Stata*, and in **Table 3** only for *Stata*

19. Answer the research question, "Is the apparent relationship between consistent condom use and existing syphilis explained by the potential confounding effects of age at first sexual intercourse?"
    a. View for yourself (but not to be included in the table) the stratum-specific odds and risk ratios
    b. Determine the adjusted odds ratio using the logistic function of *Stata* and adjusted risk ratio using the Poisson regression function of *Stata*.
        i. Include the 95% confidence intervals.
        ii. To what extent has cluster sampling increased the variances of the adjusted odds ratio and risk ratio above what they would have been if we had used simple random sampling?
    c. Present in **Table 3**

20. Answer the research question, "Is the apparent relationship between consistent condom use and existing HIV infection explained by the potential confounding effects of age at first sexual intercourse?"
    a. View for yourself (but not to be included in the table) the stratum-specific odds and risk ratios
    b. Determine the adjusted odds ratio using the logistic function of *Stata* and adjusted risk ratio using the Poisson regression function of *Stata*.
        i. Include the 95% confidence intervals.
        ii. To what extent has cluster sampling increased the variances of the adjusted odds ratio and risk ratio above what they would have been if we had used simple random sampling?
    c. Present in **Table 3**

21. Answer the research question, "Is the apparent relationship between consistent condom use and existing syphilis explained by the potential confounding or intermediate effects of recent sexual partner being a casual partner?"
    a. View for yourself (but not to be included in the table) the stratum-specific odds and risk ratios
    b. If appropriate, determine the adjusted odds ratio using the logistic function of *Stata* and adjusted risk ratio using the Poisson regression function of *Stata*.
        i. Include the 95% confidence intervals.
        ii. To what extent has cluster sampling increased the variances of the adjusted odds ratio and risk ratio above what they would have been if we had used simple random sampling?
    c. Present in **Table 3**

22. Answer the research question, "Is the apparent relationship between consistent condom use and the prevalence of HIV infection explained by the potential confounding or intermediate effects of recent sexual partner being a casual partner?"
    a. View for yourself (but not to be included in the table) the stratum-specific odds and risk ratios
    b. If appropriate, determine the adjusted odds ratio using the logistic function of *Stata* and adjusted risk ratio using the Poisson regression function of *Stata*.
        i. Include the 95% confidence intervals.
        ii. To what extent has cluster sampling increased the variances of the adjusted odds ratio and risk ratio above what they would have been if we had used simple random sampling?
    c. Present in **Table 3**

23. Answer the research question, "After controlling for the potential effects of recent sex partner being a casual partner and age at first sexual intercourse, are sexually-active women who reported consistent condom use more likely to be infected with syphilis than women who did not report that their most recent partner consistently used a condom?"

a. If appropriate, determine the adjusted odds ratio using the logistic function of *Stata* and adjusted risk ratio using the Poisson regression function of *Stata*.
    i. Include the 95% confidence intervals.
    ii. To what extent has cluster sampling increased the variances of the adjusted odds ratio and risk ratio above what they would have been if we had used simple random sampling?
b. Present in **Table 3**

24. Answer the research question, "After controlling for the potential effects of recent sex partner being a casual partner and age at first sexual intercourse, are sexually-active women who reported consistent condom use more likely to be infected with HIV than women who did not report that their most recent partner consistently used a condom?"
    a. View for yourself (but not to be included in the table) the stratum-specific odds and risk ratios
    b. If appropriate, determine the adjusted odds ratio using the logistic function of *Stata* and adjusted risk ratio using the Poisson regression function of *Stata*.
        i. Include the 95% confidence intervals.
        ii. To what extent has cluster sampling increased the variances of the adjusted odds ratio and risk ratio above what they would have been if we had used simple random sampling?
    c. Present in **Table 3**

25. Answer the research question, "Are sexually active women with syphilis more likely to be infected with HIV than women with no syphilis?"
    a. Determine the crude prevalence ratio and crude odds ratio with both the *Complex Sample* module of *Epi Info* and with the survey modules of *Stata*.
        i. Include the 95% confidence interval
        ii. To what extent has cluster sampling increased the variances of the odds ratio and risk ratio above what they would have been if we had used simple random sampling?
    b. Present in **Table 2** for *Epi Info* and *Stata*, and in **Table 3** only for *Stata*

26. Answer the research question, "After controlling for the potential confounding effect of "casual sex partner" and "consistent condom," are sexually active women with syphilis more likely to have HIV infection?"
    a. View for yourself (but not to be included in the table) the stratum-specific odds and risk ratios
    b. If appropriate, determine the adjusted odds ratio using the logistic function of *Stata* and adjusted risk ratio using the Poisson regression function of *Stata*.
        i. Include the 95% confidence intervals.
        ii. To what extent has cluster sampling increased the variances of the adjusted odds ratio and risk ratio above what they would have been if we had used simple random sampling?
    c. Present in **Table 3**

References

The references are found as PDF files at:
http://www.ph.ucla.edu/epi/rapidsurveys/RScourse/RSreadings.html.

1. Boisier P, Ouwe Missi Oukem-Boyer ON, Amadou Hamidou A, Sidikou F, Ibrahim ML, Elhaj Mahamane A, Mamadou S, Sanda Aksenenkova T, Hama Modibo B, Chanteau S, Sani A, Louboutin-Croc JP. Nationwide HIV prevalence survey in general population in Niger. *Tropical Medicine and International Health* 9(11), 1161-1166., 2004 (**Reference 19**).
2. Frerichs RR, Silarug N, Eskes N, Pagcharoenpol P, Rodklai A, Thangsupachai S, Wongba C. Saliva-based HIV-antibody testing in Thailand. *AIDS* 8(7), 885-894, 1994 (**Reference 20**).
3. West B, Walraven G, Morison L, Brouwers J, and Bailey R. Performance of the rapid plasma reagin and the rapid syphilis screening tests in the diagnosis of syphilis in field conditions in rural Africa. *Sexual Transmitted Infections* 78, 282-285, 2002 (**Reference 21**).

Population Data to Be Enter into NIGERHIV.csf

Population Data to be Entered (*.csf)			
Student (1)		Student (2)	
ComC_Dist31	3143	ComC_Dist51	3149
ComC_Dist32	3095	ComC_Dist52	3257
ComC_Dist33	3070	ComC_Dist53	3173
ComC_Dist34	3126	ComC_Dist54	3141
ComC_Dist35	3129	ComC_Dist55	3095
ComC_Dist36	3144	ComC_Dist56	3101
ComC_Dist37	3100	ComC_Dist57	3046
ComC_Dist38	3161	ComC_Dist58	3138
ComC_Dist39	3068	ComC_Dist59	3103
ComC_Dist40	3072	ComC_Dist60	3103
ComC_Dist41	3029	ComC_Dist61	3044
ComC_Dist42	3173	ComC_Dist62	3067
ComC_Dist43	3031	ComC_Dist63	3144
ComC_Dist44	3117	ComC_Dist64	3023
ComC_Dist45	3037	ComC_Dist65	2981
ComC_Dist46	3180	ComC_Dist66	3106
ComC_Dist47	3104	ComC_Dist67	3073
ComC_Dist48	3071	ComC_Dist68	3045
ComC_Dist49	3100	ComC_Dist69	3145
ComC_Dist50	3044	ComC_Dist70	1527