RAPID SURVEYS

This applied on-line course was taught like a workshop with a general outline (shown here), but no daily beginning or ending times. It was taught at UCLA as EPI 418 by Professor Frerichs for the final time in the Spring Quarter, 2008.

I. Overview of Research Methods and Issues

A. Introduction

1. Previously taught as EPI 418 Rapid Epidemiological Surveys in Developing Countries
   a. 1988 through 2008
   b. Four Units, Tuesdays and Thursdays, from 4 to 6 pm
2. Instructor
   a. Ralph R. Frerichs, D.V.M., Dr.P.H., Professor Emeritus, UCLA Department of Epidemiology
3. Opening remarks
4. Introduction of instructor and students
   a. Education and current position
   b. Experience with research and services
   c. Expectations of course
5. Objectives of course
6. Guide to course outline and readings
7. Websites
   a. Rapid Survey course
      (1) http://www.ph.ucla.edu/epi/rapidssurveys/RScourse/
   b. Rapid Surveys
      (1) Contains rapid survey references and PDF files from 1965 to the present
      (2) http://www.ph.ucla.edu/epi/rapidssurvey.html
     
     Discussion of Reference 1 (also the reference for Problem 1)

8. Use of rapid surveys
   a. Describe health problems, knowledge, attitudes and practices in populations
   b. Determine possible causative factors
   c. Evaluate programs and projects
9. Policy issues
   a. Cost-benefit analysis and cost-effectiveness analysis
   b. Efficacy, effectiveness and efficiency
   c. Forms of evaluation
      (1) input-output
      (2) input-effect
      (3) input-impact
End of Lecture 1

d. Evaluate efficiency
   (1) cost in relation to community effectiveness
   (2) estimate community effectiveness with formula
   (3) estimate efficacy with well-conducted research studies
   (4) estimate detection sensitivity
   (5) estimate compliance of health workers and patients or participants
   (6) estimate coverage
e. Do not oversell a program prior to evaluation

10. Preventing disabled years or premature death years
    a. Social values and mortality

11. Mortality indicators for cost-effectiveness analysis
    a. Mortality rates
    b. Years of potential life lost

12. Incidence-disability-mortality indicators for cost-effectiveness analysis
    a. Healthy life lost
    b. Disability-adjusted life years (DALYs)
       (1) premature death component
       (2) incident disease with disability component
       (3) anticipated effects of control strategies on DALYs

End of Lecture 2

Discussion of Reference 2

13. Rapid microcomputer-assisted surveys
    a. Components
    b. Hardware and software

B. Questions

1. Who is to be surveyed?
   a. Target population
   b. Study population
   c. Potential bias
2. What unit is to be sampled?
   a. Households
   b. Individuals
   c. Episodes, events or beliefs
3. How much error is acceptable?
   a. Truth
   b. Systematic bias
   c. Accuracy
4. How valuable is the information?

C. Planning a survey

1. Frequency of surveys
2. Required information and cost  
3. Standardized measurements  
4. Instruments for obtaining measurements  
5. Plausible values for measurements (range and logic check)  

**End of Lecture 3**  

**Discussion of Reference 3**  

6. Determine sampling strategy and design  
   a. Simple random sampling (SRS)  
   b. Systematic sampling (SS)  
   c. Stratified random sampling  
   d. Simple cluster sampling  
   e. Two stage cluster sampling  
      (1) Procedure one  
         (a) First stage: probability proportionate to size (PPS)  
         (b) Second stage: constant number selected by SRS or SS  
      (2) Procedure two  
         (a) First stage: PPS  
         (b) Second stage: constant number selected by probability sampling with quota  

**End of Lecture 4**  

**Discussion of Reference 4**  

D. Interview surveys  
   1. General comments  
   2. Standardization  
   3. Recognition of illness  
      a. Examination  
         (1) Health professional  
         (2) Lay person  
      b. Interview  
         (1) Health professional  
         (2) Lay person  
   4. Recognition of events  
      a. Truth  
      b. Reported  

E. Hardware and Software  
   1. Portable computers  
   2. Portable printers  
   3. Software  
      a. Word-processing  
      b. Management forms
c. Sample management
d. Spreadsheet
e. Statistical software
   (1) For frequencies, univariate and bivariate analyses of cluster surveys
       (a) Epi Info
   (2) For multivariate analyses of cluster surveys
       (a) Stata, Sudaan

F. Rapid survey

1. Planning
   a. Formulate the study objectives
   b. State the aims
   c. Create or select the variables
   d. Feasibility study
   e. Small pilot study

2. Organizing and conducting the survey
   a. Get local approval
   b. Involve local officials
   c. Language problems
   d. Introduction to subjects
   e. Supervision
   f. Preliminary analysis in field
   g. Present to local officials and guests
   h. Final analysis
   i. Final report

End of Lecture 5

3. Example (Myanmar)
   a. Explain survey methods
   b. Train the field staff
   c. Meet with the regional public health officials
   d. Identify 25-30 communities to be visited, each containing one or more clusters
   e. Travel to selected communities (i.e., location of clusters)
   f. Meet with community leaders
   g. Select constant number of persons (or households) per cluster
   h. In each selected cluster...
      (1) select Random Start household (HH)
          (a) if eligible, conduct interview/examinations
      (2) go to next nearest HH
          (a) if eligible, conduct interview/examinations
      (3) continue until set number (e.g., 10) is interviewed or examined
          (a) typically the total survey has 30 or more clusters
          (b) with 10 persons being interviewed or examined per cluster, the sample size for the survey would be 300
   i. Every afternoon...
      (1) Enter data into portable computer
      (2) Review data for errors or omissions
(3) Determine if interviewer needs to return to the field
j. At the end of the week before leaving the field...
(1) Do frequencies of all variables and cross-tabulations of selected variables
(2) Print selected tables and graphs with portable computer and printer
(3) Share selected findings with local health officials and community leaders
k. At central headquarters...
(1) Present main findings with graphs and some tables to the national staff and other interested persons
(2) Complete the final analysis
(3) Prepare written report for national distribution that includes graphs of major findings and tables of all results
l. Time frame
(1) collect data for 4-5 days
(2) do analysis and write final report in 6 to 9 days

G. Interview training

1. Types of interviews
   a. Confidential
   b. Non-confidential
2. Approach to interviews
3. Use of role playing
   a. Example
   b. Discussion

End of Lecture 6

II. Statistical Methods and Issues For Rapid Surveys

A. General Notions

1. More on units
   a. Sampling and elementary units
   b. Enumeration rule
2. Value of alternative sources of information
   a. Opinion of expert
   b. Consensus of experts
      (1) Delphi procedure
   c. Synthetic estimation
   d. Sample of population
   e. Total count of population

Discussion of Reference 5

3. Data
   a. Types of data
      (1) Binomial
      (2) Equal interval
      (3) Ratio of two variables
b. Population and sample means
c. Derivations from data of mean and variance of sample mean
   (1) Equal interval
   (2) Binomial
   (3) Ratio estimator
d. Derivation of confidence interval
e. Data and action

**End of Lecture 7**

4. Variability and bias
   a. A survey of 20
      (1) precision or accuracy
   b. Concept of...
      (1) bias
      (2) accuracy
      (3) precision
c. Value of statistical tests
d. Measuring precision
   (1) variance and standard error of the sample mean
   (2) distribution of all possible samples
      (a) change in scale from outcome values to standard error values
e. Confidence interval
   (1) playing...
      (a) darts with variable-sized heads
      (b) confidence intervals
   (2) driving confidence intervals
   (3) distribution of confidence intervals of all possible samples
   (4) notions of probability versus confidence

**End of Lecture 8**

**Discussion of Reference 6**

B. Simple Random Sampling

1. Introduction
   a. Intravenous drug addict example
      (1) Draw 3 drug addicts from a population of 9 drug addicts
      (2) True average value in population
      (3) All possible samples of 3 from population of 9
         (a) Sampling with replacement
         (b) Sampling without replacement
      (4) Finite population correction (FPC) term
   2. Equal interval variable (1)
      a. Total IV injections during prior week
         (1) True values in population
         (2) Mean and standard error of all possible samples
         (3) 95% confidence interval of all possible samples
3. Equal interval variable (2)
   a. Shared IV injections during prior week
      (1) True values in population
      (2) Mean and standard error of all possible samples
      (3) 95% confidence interval of all possible samples

End of Lecture 9a

4. Binomial variable
   a. Infected with HIV
      (1) True values in population
      (2) Proportion and standard error of all possible samples
      (3) 95% confidence interval of all possible samples

5. Ratio estimator
   a. Ratio of shared IV injections to total IV injections during prior week
      (1) True values in population
      (2) Ratio estimator and standard error of all possible samples
      (3) 95% confidence interval of all possible samples

C. Samples and elements
   1. Sample units
   2. Elementary units

D. Survey of smoking behavior (example)
   1. Simple random sample survey (54% smokers)
      a. Sample of 90 persons from population of 3000 persons
         (1) Relation of t value to z value
         (2) 95% confidence interval of 100 of all possible samples

End of Lecture 9b

2. One stage cluster survey (54% smokers)
   a. Sample of 30 households (HHs) from population of 1000 HHs
      (1) Relation of t value to z value
      (2) Persons per HH with average size of 3 persons/HH
      (3) Two scenarios
         (a) HH size is not associated with percentage of smokers per HHs
         (b) HH size is associated with percentage of smokers per HHs
      (4) Variability among and within clusters (i.e., HHs)
         (a) variance of a binomial variable
         (b) variability among clusters
            i) high
            ii) medium
            iii) low
      (5) Population distortion and possible bias in ratio estimator
         (a) Distorted population (but still with 54% smokers)
            i) high variability of smoking among (i.e., between) HHs
            ii) smoking status varies dramatically with HH size
         (b) High variability but only slight bias in ratio estimator
End of Lecture 9c

Discussion of Reference 7

E. Equal Probability of Selection

1. Introduction
   a. Sampling 3 addicts from a population of 9 addicts
   b. Sampling of 20 HHs from population of 10,000 HHs

2. Equal probability of selection method (EPSEM) sampling

3. Two-stage cluster sampling
   a. with equal selection probabilities
      (1) unequal sized clusters
         (a) equal fraction at second stage
      (2) equal sized clusters
         (a) equal number at second stage
   b. with unequal selection probabilities
      (1) unequal sized clusters
         (a) equal number at second stage

4. Probability proportionate to size (PPS) Sampling
   a. with equal selection probabilities
      (1) unequal sized clusters
         (a) equal number at second stage
   b. Method
      (1) natural order and systematic sampling
      (2) random order with systematic sampling
      (3) cluster-specific random PPS sampling

5. Sampling and elementary units
   a. Representative
   b. Not representative

End of Lecture 10

Discussion of Reference 8

F. Mean and Variance of Cluster Sampling

1. Sampling unit same as elementary unit
   a. Sampling of persons
   b. When all clusters are the same size
      (1) Mean
         (a) Binomial
         (b) Equal interval
      (2) Variance
         (a) Binomial
            i) the simple formula
         (b) Equal interval
            i) the simple formula
      (3) 95% confidence interval
(a) Binomial
(b) Equal interval

2. Sampling unit different from elementary unit
   a. Sampling of households
   b. Clusters have constant HHs but varied elementary units
      (1) Analysis of total group
          (a) All elementary units in each cluster
          (b) Ratio estimator
              i) sample ratio estimator (r) acting as a proportion
                 a) mean
                 b) variance
                    1) the complex formula
                 c) 95% confidence interval

End of Lecture 11a

Discussion of Reference 9

ii) sample ratio estimator (r) acting as a mean
    a) mean
    b) variance
       1) the complex formula
    c) 95% confidence interval

(2) Analysis of subgroups
   (a) Incorrect analysis
      i) use simple variance formula
   (b) Correct analysis
      i) use complex variance formula

3. Conclusions
   a. Simple variance formula is only correct in certain circumstances
   b. For persons doing rapid surveys...
      (1) Use complex variance formula for all analyses
   c. Sampling at second stage after PPS sampling at first stage
      (1) If only one eligible person is likely to be in a HH
          (a) sample constant number of eligible persons per cluster, or
          (b) sample constant number of eligible HHs per cluster
      (2) If HHs likely contain more than one eligible person
          (a) sample constant number of occupied HHs per cluster

End of Lecture 11b

Discussion of Reference 10

G. Cluster Sampling for Common Events

1. Introduction
   a. Steps for doing a survey
   b. Potential biases
2. Steps for doing a EPI/WHO style two-stage cluster survey
a. First stage
   (1) Identify young child population
   (2) Identify young child population that can be sampled
   (3) Select sampling units from sampling frame
b. Second stage
   (1) Select random start household
   (2) Determine if someone is home at random-start HH
   (3) With permission, interview mother or other knowledgeable respondent
   (4) Continue to next nearest HH
   (5) Repeat process until quota of seven eligible children have been studied in each cluster

3. Sample size determination
   a. Answer three questions
      (1) Best estimate of true proportion in study population?
      (2) How precise should the estimate be?
      (3) How confident do you need to be that a derived interval brackets the true value in the study population?
   b. Use Csurvey program to derive appropriate sample size
      (1) Answer to three questions
      (2) Design effect
      (3) Intraclass correlation coefficient
      (4) Maximum standard error

4. Sample size and decision-making
5. Evaluation of WHO-style surveys

End of Lecture 12

Discussion of Reference 11

H. Other Uses of Cluster Sampling

1. Measure changes in level of an attribute
   a. Hypothesis testing versus interval estimation
      (1) Type I and II errors
      (2) Power of the test
   b. Measuring attribute at same time in different sites
   c. Measuring attribute at different times in same sites
2. Assess the occurrence of uncommon events
   a. Not recommended
3. Other applications
   a. Measure knowledge, attitudes and practices (KAP)
   b. Deriving controls for a population-based case-control study

I. Cluster Sampling for Rare Events

1. May be requested by governments or non-governmental organizations (NGOs)
   a. Such surveys tend to be very costly
   b. Examples
2. Approach to such survey requests favored by R. R. Frerichs
a. Use Csurvey program
b. Example - HIV prevalence survey
   (1) Interval estimation
   (2) Hypothesis testing

End of Lecture 13

J. Use of design effect (and roh)
   1. Introduction
      a. Terms
         (1) deff = design effect
         (2) roh = intraclass correlation coefficient
      b. Most important use is for planning surveys
      c. Secondary use is for management

Revisit of Reference 4

2. Equation for variance of sample mean
   a. Deviation of observed from expected
   b. Variation among clusters
      (1) High variation with design effect of 4.0
      (2) Low variation with design effect of 1.0
         (a) identical to variation if SRS
      (3) Very high variation with design effect of 10.7

3. Values of design effect in two-stage cluster surveys
   a. Variables in maternal and child health survey
   b. Coverage as assessed in immunization surveys
   c. Infectious disease incidence surveys

4. Uses of design effect for management
   a. For health care variables
   b. For disease variables

End of Lecture 14

Discussion of Reference 12

K. Improving Precision of Rapid Surveys

1. Theory
   a. Most statistical tests assume data are drawn from some underlying population
      and are independent observations
   b. Homogeneity of findings within clusters biases statistical tests

2. p-values and confidence intervals
   a. Underestimated with conventional statistical software
   b. Implies statistical significance when there is none

3. Actual field experience
   a. Comparison of two sampling techniques
      (1) EPI-style
         (a) two-stage cluster with PPS at first stage and constant number with
random start and next nearest neighbor at second stage
(2) SRS-style
   (a) two-stage cluster with PPS at first stage and constant number with
       simple random sample at second stage
4. Steps to consider for improving precision
   a. Estimate roh based on prior surveys or review of the literature
   b. Strategy for attributes...
      (1) with small values of roh
      (2) with large values of roh
   c. Example with multiple independent selections of random start households at
      second stage

End of Lecture 15a

Discussion of References 13

L. Improving the Accuracy of Rapid Surveys

1. Difference between accuracy and precision
   a. Precision
   b. Bias
   c. Accuracy
2. Sources of potential bias
   a. Probability sampling with quotas
   b. Design bias
   c. Measurement bias
   d. Data processing or analysis bias
   e. Selection bias
   f. Bias due to EPI/WHO-style of cluster sampling
3. Non-response errors with two-stage cluster surveys

Discussion of References 14

4. Modifications to improve accuracy
   a. Not at home problem
   b. Modifications for dispersed rural areas
      (1) Use of school children
   c. Modifications for urban areas
      (1) First stage sampling of clusters
      (2) Second stage sampling of selected clusters
         (a) Two options

End of Lecture 15b

Discussion of References 15 and 16

M. Example of Rapid Survey in an Urban Area

1. Young Child Immunization Survey
a. Three inner-city zip code areas
b. Los Angeles, California

2. How the survey was conducted
a. At home parents of children, aged 2-3 years
b. 271 face-to-face interviews

3. Sampling scheme for two-stage cluster survey
a. First stage PPS sampling of clusters
b. Second stage SR sampling of persons

4. Create map for each cluster

5. Phases
   a. Enumeration phase
      (1) use volunteers and immunization program staff
   b. Interview phase
      (1) use UCLA students and immunization staff

6. Special research component
   a. Collection of saliva specimen for determination of measles immunity

7. Participation - two methods
   a. Denominator is sampled children
   b. Denominator is at-home children

8. Telephone status of participating households

**End of Lecture 16**

**III. Closing comments**

A. Completion of course – what was gained?

1. Appreciate the need for rapid surveys
   a. Descriptive (cross-sectional) studies
   b. Etiologic studies
   c. Effectiveness studies

2. Understand the theory of...
   a. sampling
   b. rapid surveys
   c. assessment for decision-making

3. Acquired practical skills to...
   a. do the field work for rapid surveys
   b. analyze data from rapid surveys

B. Final Group Problem

1. What will be gained - experience
   a. Objectives
   b. Methods
   c. Results
   d. Discussion and recommendations

2. Able to present analysis of survey findings in terms understandable to a non-technical audience

3. Present in-class during three-hours scheduled for the final examination
a. Monday evening
b. Two independent presentations by Groups A and B

C. Concluding remarks

1. Intent and attempt of rapid surveys

End of Lecture 17

GRADING

Individual Problem 1 – Graphics ............................................ 15%
Team and Individual Problem 2 – Data Analysis ..................................... 35%
Final Group Problem – Rapid Survey ........................................ 50%

REFERENCES


**SUPPLEMENTARY REFERENCES**


**SUPPLEMENTARY REFERENCES FOR PROBLEM 2**


**HISTORICAL REFERENCES FOR FINAL GROUP PROBLEM**

