

Jason & Glenwick, eds.
Methodological Approaches to Community-Based Research
Chapter 11/Web Appendix guide

Web Appendix Guide for Chapter 11/ Epidemiologic Approaches

In order to execute the calculations proposed in this chapter, the reader will need four files:

1. This main appendix file of instructions (Jason.Appendix.Main.docx)
2. The SAS program file (Jason.Appendix.SASProgram.sas). This program reads in the SAS data file, performs some basic statistical frequencies and performs the analyses to estimate overall prevalence as well as prevalence by gender. The reader is instructed to perform the calculations by race, age and SES.
3. The SAS data file (JasonAppendixSASDATA.sas7bdat). This is a SAS system datafile. Note that this file should have no dots (periods) in the name, except for the extension '.sas7bdat'.
4. The Excel file (Jason.Appendix.Excel.xls). This file takes frequencies from the SAS output and calculates the prevalence, standard error of prevalence, and 95% confidence interval for the prevalence of CFS, ICF, and CF-explained.

Readers should refer to Shrout, P. E., & Newman, S. C. (1989). Design of two-phase prevalence surveys of rare disorders. *Biometrics*, 45, 549–555, which is the basis for the calculations in this Web Appendix. Sufficient information will be given in this Web Appendix to perform all the calculations, but the paper provides the setting for the two-stage screening design, the background and the full notation for the calculations.

The following steps should be done by the reader to complete the calculations:

1. Store the SAS data file in a specific folder. The pathname for this folder is required in the SAS program file.
2. Open the SAS program. The first statement in the SAS program is the following:

LIBNAME PREV 'D:\F5 CFS\EPI CHAPTER\';

In this statement, the part that says **D:\F5 CFS\EPI CHAPTER** should be replaced by the **pathname for the folder where the SAS data file is stored.**

3. Run the SAS program.

Once the pathway for the SAS dataset has been specified in the SAS program, the program will format the data and run the appropriate frequencies that are needed to calculate the prevalence, standard error of prevalence, and 95% confidence interval for prevalence.

The following variables in the SAS dataset are needed to obtain the intermediate frequencies that are required:

SCREEN: formatted as 0=screen negative in Phase 1, 1=screen positive in Phase 1

WORKUP: formatted as 0=not worked up in Phase 2, 1=worked up in Phase 2

CFSSTAT: diagnosis after workup, formatted as 1=CFS, 2= ICF, 3= CFS-explained, 4=control

GENDER: formatted as 1=female, 2=male

RACE: formatted as 1=Black, 2=White, 3= Latino, 4=other

AGECAT: formatted as 18-29, 30-39, 40-49, 50-59, 60+

SES: formatted as unskilled/semiskilled, skilled, professional

The SAS program may be run by clicking on the Run icon in SAS. After the program has run, go to the OUTPUT folder where the results of the calculations are given. This will be an electronic folder that is populated after the SAS program runs.

4. Extract the information from the OUTPUT folder and enter into the Excel file.

The calculations will now be illustrated for the sample as a whole and for GENDER. The numbers to do the calculations will be in the OUTPUT folder.

The reader may then perform the calculations for RACE, AGECAT and SES to verify the prevalence rates given in Table 11.2 of Chapter 11.

Calculations for CFS prevalence, overall (* means multiply, / means divide)

a) N is the total number screened, and is 18668 for the overall sample.

a) $PI = \text{proportion of screen positives in Phase 1} = 408/18668 = .0219$. A frequency on the variable SCREEN will give PI.

b) $F1 = \text{proportion of screen positives worked up in Phase 2} = 166/408 = .4069$.

$F2 = \text{proportion of screen negatives worked up in Phase 2} = 47/18260 = .0026$.

A crosstabulation of SCREEN*WORKUP will give these numbers.

c) $L1 = \text{proportion of worked up screen positives that have CFS} = 32/166 = .1928$

$L2 = \text{proportion of worked up screen negatives that have CFS} = 0/47 = 0$

A crosstabulation of CFSSTAT*SCREEN will give these numbers

d) Take the values N, PI, F1, F2, L1, L2 and put them into the spreadsheet in the first row and in the appropriate column.

e) Prevalence = $L1*PI + L2*(1-PI) = .1928*.0219 + 0*.9781 = .00422$

f) Variance of Prevalence =

$$[PI*L1*(1-L1)/F1 + (1-PI)*L2*(1-L2)/F2 + pi*(1-PI)*(L1-L2)*(L1-L2)] / N$$

Where N = total sample size of 18668

Substituting the appropriate values, the variance of the prevalence is

$$[.0219*(.1928)*(.8072)/.4069 + .9781*0*1/.0025 + .0219*.9781*.1928*.1928] / 18668$$

$$= .009172 / 18668 = .0000004913$$

g) Standard error of prevalence = square root of the variance = .00070

h) 95% confidence interval for the prevalence is

prevalence +/- 1.96*standard error of prevalence =

$$.00422 \pm 1.96*.00070 = .00285, .00560$$

i) To summarize, prevalence (se) = .00422 (.00070), 95% confidence interval is (.00285, .00560).

j) In paper, rates are per 100,000, so multiplying the above numbers by 100,000

prevalence (se) = 422 (70), 95% confidence interval is (285, 560).

Calculations for CFS prevalence, by gender

SAS program gives the above statistics by gender.

For Female, N=10507, P1=.0279, F1=.4198, F2=.0023, L1=.1870, L2=0. Calculations in Excel give

prevalence (se) = .00522 (.00103), 95% confidence interval is (.00321, .00723).

For Male, N=8110, P1=.0139, F1=.3805, F2=.0030, L1=.2093, L2=0. Calculations in Excel give prevalence (se) = .00291 (.00091), 95% confidence interval is (.00113, .00468).

The reader may calculate prevalence by race, agecat and ses.

Similar calculations may also be done for CFS and CF-Explained, since these also are outputs from the SAS program using the CFSSTAT variable.