

## Estimating Prevalence When a Subset of the Sample Requires Further Evaluation

David C. Hoaglin<sup>1</sup> and K. P. Srinath<sup>2</sup>  
 Abt Associates Inc., Cambridge, MA<sup>1</sup>  
 Abt Associates Inc., Bethesda, MD<sup>2</sup>

### Abstract

Population-based estimates of the prevalence of some health conditions require that a subset of the survey sample have a clinical evaluation, in addition to an initial personal interview. Chronic fatigue syndrome (CFS) provides a motivating example: an interview can elicit whether respondents have the symptoms that are part of the definition of CFS, as well as whether they have certain exclusionary conditions, but the actual diagnosis of CFS depends on laboratory data and the results of a physical examination and other tests. Thus, all subjects with sufficient symptoms (and no exclusions) are eligible for clinical evaluation, and other subsets of subjects may be selected for comparison. For types of fatiguing illness whose definition involves clinical data, the paper examines two approaches for estimating standard errors of prevalence estimates.

**Keywords:** Survey, Variance estimation, Chronic fatigue syndrome

### 1. Introduction

To develop population-based estimates of the prevalence of some relatively rare health conditions, one approach uses random-digit dialing to identify households. A household respondent is asked to enumerate the members of the household and provide information on their demographic characteristics and relevant aspects of their health. On the basis of that information, persons with certain characteristics are selected for detailed interviews about their health. The information that they provide permits them to be classified in more detail, relative to the condition under study, in part by taking into account various other, exclusionary, conditions. For the sorts of conditions with which we are concerned, reaching a diagnosis requires a clinical evaluation. Thus, subjects who have not been excluded are invited, perhaps on a sampling basis, to undergo clinical evaluation. The Georgia Telephone Survey of Chronic Fatigue Syndrome, described in Section 2, has this structure.

Such a survey yields two main sampling weights: one for the persons who complete detailed interviews and the other for the subset who complete clinical evaluations. For health conditions whose definitions

involve only data from the detailed interviews (and the screening interview with the household respondent), the prevalence estimates are weighted proportions, using the interview weights, and calculation of their standard errors is straightforward. When data from the clinical evaluations are involved, the prevalence estimates are still straightforward, but the standard errors should take into account variability from sampling persons for clinical evaluations. Section 3 describes a simplified approach, based on a composite dataset with an additional set of weights. To take into account more sources of variability, Section 4 builds on results from two-phase sampling. Section 5 compares the standard errors produced by these two approaches for one key prevalence estimate; it also provides some concluding discussion.

### 2. Georgia Telephone Survey of CFS

As part of its program of research on chronic fatigue syndrome (CFS), the Centers for Disease Control and Prevention (CDC) undertook a telephone survey of CFS and chronic unwellness in three areas of Georgia (Reeves et al. 2006). The study used random-digit dialing to contact a random sample of households (with telephones) in three strata (Metropolitan, Urban, and Rural—which together did not cover the state of Georgia) between September 2004 and June 2005. The interviewers asked to speak with the member of the household who knew the most about the health of the family. The screening questionnaire then asked the respondent to enumerate the members of the household aged 18 years and older and to provide information on age, gender, race/ethnicity, and presence or absence of fatigue, unrefreshing sleep, difficulty thinking or concentrating, and pain for each household member. All persons aged 18 to 59 reported as having prolonged fatigue (i.e., severe fatigue, extreme tiredness, or exhaustion lasting one to five months) were subsequently asked to complete detailed telephone interviews. A subsample of adults reported as having unwellness (that is, problems with memory or concentration, unrefreshing sleep, or pain for one month or longer) was also selected for detailed interviews, as was a subsample of “well” adults (individuals reporting no fatigue and no problems with memory or concentration, unrefreshing sleep, or pain for one month or longer).

The sample of 105,000 telephone numbers (35,000 per stratum) yielded completed screening interviews with 10,837 households. In these households a total of 19,381 individuals were enumerated, of whom 3,425 were reported as having prolonged fatigue, 5,122 were reported as unwell, and 10,834 were reported as well. From the 5,122 individuals who were reported to be unwell, 2,134 were selected for detailed interviews. Similarly, from the 10,834 individuals who were reported to be well, 3,113 were selected.

In these three samples, 2,438 fatigued individuals, 1,429 who were unwell without fatigue, and 1,756 who were well completed detailed interviews. Each of these 5,623 individuals received a sampling weight that reflected the probability that the person was selected for a detailed interview. The sampling weights also incorporated adjustments for nonresponse (e.g., interviews that could not be conducted and households for which screening questionnaires were not completed) and an adjustment that used information on interruptions in telephone service to make allowance for persons in households without telephones (Frankel et al. 2003). An iterative process of poststratification brought the weighted total from each stratum into agreement with population control totals from the 2000 Census on two categories of race (Black, White/Other) and 18 combined categories of sex and age (nine categories of age for each sex).

On the basis of the data from their detailed interviews, subjects were assigned to one of three sample categories: CFS-like illness (907 subjects), chronically unwell (2,633), and well (2,083). Eligibility for a clinical evaluation required the absence of certain medical and psychiatric conditions. After these exclusions, 469 CFS-like, 1,763 chronically unwell, and 1,782 well subjects remained eligible.

The 469 CFS-like persons were asked to undergo clinical evaluations. Clinical evaluations were actually completed on 292 of these persons; and 84 of them met the definition of CFS. Of the other 208, 66 were classified as having insufficient symptoms or fatigue (ISF), 141 had medical or psychiatric exclusions (identified in the clinical evaluation), and 1 could not be classified.

From the 1,763 eligible chronically unwell persons, a subsample of 505 was selected for clinical evaluation. Among the 268 who completed clinical evaluations, the distribution of classifications was 26 CFS, 126 ISF, 26 well, 89 excluded, and 1 not classified.

A well subject was selected for clinical evaluation if s/he matched a clinically evaluated CFS-like subject

on the basis of geographic stratum (Metropolitan, Urban, and Rural), sex, race, Hispanic/non-Hispanic ethnicity, and age (within three years). Among the 223 well subjects who completed clinical evaluations, the classifications were 3 CFS, 72 ISF, 98 well, and 50 excluded.

The CFS-like and chronically unwell subjects who completed clinical evaluations each received a clinical weight, which incorporated a further adjustment for nonresponse (i.e., clinical evaluations that could not be conducted—on 177 CFS-like subjects and 237 chronically unwell subjects). It also reflected the probability that the chronically unwell person was selected for the clinical-evaluation subsample. Because well subjects were selected for clinical evaluation only as a result of being matched to CFS-like subjects, those well subjects who completed clinical evaluations did not have their own clinical-evaluation weight. The data from their clinical evaluations were not used in calculating prevalence estimates.

### 3. Simplified Approach

For a particular category of fatiguing illness, the estimate of prevalence is the weighted percentage of the sample in that category. For categories other than CFS and ISF (e.g., chronically unwell with no fatigue), the estimates are derived from the detailed interview data, using the interview weight.

To estimate the prevalence of CFS (or ISF) and the corresponding standard error, we developed a composite dataset, which involved data from both the detailed interviews and the clinical evaluations. To construct the composite dataset, we began with the detailed interview dataset, which contained data for the 5,623 subjects who completed detailed interviews. We then removed the data of the 469 CFS-like subjects and the 1,763 chronically unwell subjects who were eligible for clinical evaluations. As their sampling weight in the composite dataset, the remaining 3,391 subjects (the “well” subjects and those with exclusions) retained their value of the interview weight. From the clinical evaluation dataset, we included the 292 CFS-like subjects and the 268 chronically unwell subjects who completed clinical evaluations. (That is, the CFS-like subjects who completed clinical evaluations replaced those—including themselves—who were eligible for clinical evaluations, and similarly for the chronically unwell subjects who completed clinical evaluations.) As their sampling weight in the composite dataset, these 560 subjects had their value of the clinical weight. Because of the way in which the clinical weight was

calculated, the total sampling weight of the 3,951 subjects in the composite dataset was the same as the total of the interview weight in the detailed interview dataset. The composite dataset yielded estimates of the prevalence of CFS and ISF, along with appropriate standard errors.

Table 1. Disposition of Interview Cases in the Rural Stratum of the Georgia Survey

Disposition	CFS-like	Chron. Unwell	Well	Total
Ineligible for clinical evaluation (exclusion)	<i>238</i>	<i>419</i>	<i>125</i>	<i>782</i>
Not selected for clinical evaluation	0	603	484	1,087
Selected, not completed	89	125	202	416
Selected, completed	<i>141</i>	<i>134</i>	<i>109</i>	<i>384</i>
Total	<i>468</i>	<i>1,281</i>	<i>920</i>	<i>2,669</i>

Numbers of subjects in composite dataset are in italics.

Table 1 shows the various subsets of subjects in this process for the Rural stratum of the Georgia Survey. The numbers of subjects in the subsets that make up the composite dataset are shown in italics. For this stratum the composite dataset contains  $238 + 419 + 125 + 484 + 202 + 141 + 134 + 109 = 1,852$  subjects.

In order to take into account the sampling weights and clustering of subjects within households (rare, but not absent), we used SUDAAN to calculate standard errors. The calculation used the Taylor-series method and assumed with-replacement sampling of households (the primary sampling unit). The estimated prevalence of CFS in the Rural stratum was 2.659%; and the simplified approach, based on the composite dataset, gave an estimated standard error of 0.575%.

#### 4. Alternative Approach

As an alternative approach for estimating the standard error of the prevalence estimate for CFS (or ISF), one can use results from two-phase sampling. In the particular form of two-phase sampling, the first phase provides a basis for developing the subsets of the population and estimating their relative size. (We refer to “subsets” rather than the customary “strata” to minimize confusion with the geographic strata of the

Georgia Survey. We are concerned here only with estimation within those strata.) The samples at the second phase, within the subsets, then yield subset-specific estimates. The overall estimate is the appropriate weighted average of those estimates, and its variance incorporates information from both phases. In the application to the Georgia Survey the sample from the detailed interview plays the role of the first phase, and the samples from the clinical evaluation form the second phase.

In the notation underlying the formula for the variance,  $n_h$  denotes the number of people who completed detailed interviews and fell into subset  $h$ :

$n_1$  with CFS-like illness and no exclusions,

$n_2$  chronically unwell with no exclusions, and

$n_3$  well or having exclusions;

and  $n = n_1 + n_2 + n_3$ . For person  $i$  in subset  $h$ , the interview weight is  $W_{lhi}$ . Thus, the estimated proportion of the population in subset  $h$  is

$$w_h = \frac{\sum_{i=1}^{n_h} W_{lhi}}{\sum_{h=1}^3 \sum_{i=1}^{n_h} W_{lhi}} .$$

We let  $m_h$  denote the number of people who completed clinical evaluations in subset  $h$ ,  $W_{Chi}$  denote the clinical weight of person  $i$  in subset  $h$ , and  $y_{hi}$  denote the outcome for that person (e.g.,  $y_{hi} = 1$  if the person is classified as having CFS, and  $y_{hi} = 0$  otherwise). Then the estimated prevalence in subset  $h$  is

$$p_h = \frac{\sum_{i=1}^{m_h} W_{Chi} y_{hi}}{\sum_{i=1}^{n_h} W_{lhi}}$$

and the overall estimated prevalence is

$$p = \sum_{h=1}^3 w_h p_h .$$

Finally, from Rao (1973) we have an estimate of the variance of  $p$ :

$$v(p) = \sum_{h=1}^3 \frac{w_h^2 p_h (1 - p_h)}{m_h - 1} + \frac{1}{n - 1} \sum_{h=1}^3 w_h (p_h - p)^2 .$$

The first term combines contributions from the first-phase and the second-phase sampling, and the second term arises because of the first-phase sampling.

We note that, by design,  $p_3 \equiv 0$ , mainly because persons who have exclusions cannot receive a diagnosis of CFS. Also,

$$\sum_{i=1}^{m_h} W_{Chi} = \sum_{i=1}^{n_h} W_{lhi}$$

for  $h = 1$  and  $h = 2$ .

For the estimated prevalence of CFS in the Rural stratum, this approach gave an estimated standard error of 0.612% (versus 0.575% for the simplified approach). The value of  $v(p)$  was  $3.740 \times 10^{-5}$ , and the values of the two terms were  $3.651 \times 10^{-5}$  and  $8.860 \times 10^{-7}$ , respectively.

**5. Comparisons and Conclusion**

For a broader comparison of the two approaches, Table 2 shows the results for all three strata in the Georgia Survey. The estimated prevalences in the three strata were quite similar, and the differences among their standard errors reflect the corresponding sample sizes.

In each stratum the simplified approach yielded a smaller estimated standard error than the alternative approach. The difference was noticeable in the Rural stratum but slight in the Metro and Urban strata.

In the estimates of the variance by the alternative approach, the second term makes only a minor contribution. This term, part of the first-phase variance, arises from differences among the  $p_h$  for the three subsets.

In summary, both the simplified approach (based on the composite dataset) and the alternative approach (based on two-phase sampling) yield estimates of the variance of the estimated prevalence that are approximations to the variance estimate based on the actual sampling design of the survey. In the three strata of the Georgia Survey the difference between the two approximations is relatively small.

The simplified approach does not include the variance from sampling at the first phase. The resulting underestimation will not be serious when the first-phase sample sizes are large.

Table 2. Comparison of the Two Approaches on the Three Strata of the Georgia Survey

	Stratum		
	Metro	Urban	Rural
Estimated prevalence (%)	2.552	2.482	2.659
Simplified approach			
Standard error (%)	0.846	0.667	0.575
Alternative approach			
Standard error (%)	0.857	0.679	0.612
Variance ( $\times 10^{-5}$ )	7.351	4.604	3.740
First term ( $\times 10^{-5}$ )	6.951	4.414	3.651
Second term ( $\times 10^{-5}$ )	0.399	0.190	0.089

The alternative approach does not take into account the clustering of persons within households. Thus, it also underestimates the variance, but the error will be slight when clustering occurs in only a small proportion of households.

In further work we plan to develop variance estimates that take into account all the features of designs such as that used in the Georgia Survey.

**Acknowledgment**

The authors thank Michael P. Battaglia for helpful suggestions.

**References**

Frankel, M.R., Srinath, K.P., Hoaglin, D.C., Battaglia, M.P., Smith, P.J., Wright, R.A., and Khare, M. (2003). "Adjustments for Non-telephone Bias in Random-digit-dialling Surveys," *Statistics in Medicine*, 22:1611–1626.

Rao, J.N.K. (1973). "On Double Sampling for Stratification and Analytical Surveys," *Biometrika*, 60:125–133.

Reeves, W.C., Jones, J.F., Maloney, E., Heim, C., Hoaglin, D.C., Boneva, R.S., Morrissey, M., and Devlin, R. (2006). "Prevalence of Chronic Fatigue Syndrome in Metropolitan, Urban, and Rural Georgia," unpublished manuscript, National Center for Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, GA.