The Estimation of Prevalence for the Evaluation of a Screening Program

RICHARD G. CORNELL

Department of Biostatistics, University of Michigan, Ann Arbor, Michigan 48109

The evaluation of a screening program in which only screen positives are tested with more definitive diagnostic procedures requires the estimation of prevalence so that the number of screen negatives who are actually diseased can also be estimated. Methodology for incorporating an assessment of the variability of the prevalence rate between as well as within subpopulations is presented and illustrated. The procedure described leads to a negative binomial probability distribution on the number of diseased individuals who were missed by the combined screening–diagnostic protocol.

1. INTRODUCTION

The design of health screening programs commonly leads to follow-up of individuals who are positive with more definitive diagnostic procedures but to no additional information on those who are negative. Thus in the notation presented in Table I, a and b, the frequencies of individuals who are diseased and nondiseased among screen positives, are known but for screen negatives only $c + d$, the total number, is known. Separate frequencies c and d are needed for the evaluation of such a program.

Grant (*1*) discusses the evaluation of a screening program when all the frequencies in the table are known. He also mentions the problem of obtaining c and d separately and suggests that the prevalence $(a + c)$ be estimated from a sampling study, from data from other studies, or from the "best possible guess," where such a guess might be obtained from expert opinion through use of the Delphi technique. The observed frequency a would then be subtracted from the estimated prevalence to estimate c, the number of diseased persons missed by the screening program.

The uncertainty about the prevalence is not taken into account in the evaluation methodology presented previously. For instance, Grant assumes that the national prevalence rate of 5% holds for the $n = 547$ children tested in a pure-tone hearing screening study, which leads to an estimated prevalence frequency of 27. The observed a is 20, so the estimated value of c is 7. He also observed $b = 103$ and $c + d = 424$. If the prevalence rate were 6% instead of 5% for this group, the estimated values of $a + c$ and c would be 33 and 13, respectively. Thus, a slightly higher prevalence, such as might be reasonable

TABLE I

NOTATION FOR FREQUENCIES FROM A SCREENING PROGRAM

Screening results	Diagnosis		Total
	Diseased	Nondiseased	
Positive	a	b	$a + b$
Negative	c	d	$c + d$
Total	$a + c$	$c + d$	$a + b + c + d$

for a group selected for screening, can lead to nearly doubling the estimated value of c. This in turn would change the estimated underreferral rate $(c/[n - (a + b)])$ in his example from 0.02 to 0.04.

The purpose of this note is to present methodology for incorporating the inherent variability of the prevalence rate into the evaluation of a screening program.

2. MODEL AND METHODOLOGY

To emphasize that the prevalence rate will vary about the national rate among smaller groups of people screened, denote the number of diseased persons in the group screened by the random variable X and denote the number of false negatives in the screened group by $Y = X - a$. For any sample of individuals screened from a homogeneous subpopulation, the distribution of X for the subpopulation can reasonably assumed to be Poisson. The distribution of means for such subpopulations can be represented by a gamma distribution. Compounding the conditional Poisson distribution with a gamma distribution leads to an unconditional negative binomial distribution for X, with probability function

$$g(x; \alpha, \beta) = \{\Gamma(x + \alpha)/[\Gamma(x + 1)\Gamma(\alpha)]\}[1/(1 + \beta)]^x[\beta/(1 + \beta)]^\alpha,$$
$$x = 0, 1, \ldots; \alpha, \beta > 0. \quad [1]$$

This distribution has mean $E(X) = \alpha/\beta$, which could be specified on the basis of a national rate, and variance $V(X) = E(X)[(1 + \beta)/\beta]$, which would reflect variation among as well as within subpopulations.

In order to determine the distribution of Y, assume that the distribution of a conditional upon X and p is binomial, where p represents the probability that a diseased individual will be found positive with the combination of a screening examination and follow-up with the diagnostic procedure; that is, p is the sensitivity of the combined screening–diagnostic procedure. We will assume that p is specified. This is reasonable since the uncertainty about p is usually of smaller magnitude than the uncertainty about the prevalence in the sample screened. The analysis would usually be repeated for several values of p to investigate the sensitivity of the evaluation of the screening procedure to the specification of p.

Once the screening program is carried out and a is observed, the distribution of Y, the number of undetected diseased individuals, is negative binomial with parameters α and β in [1] replaced by $\alpha^* = a + \alpha$ and $\beta^* = (p + \beta)/(1 - p)$, respectively (see Appendix). The mean of this distribution, $E(Y) = \alpha^*/\beta^*$, can be used as a point estimate of c. Percentiles of the conditional negative binomial distribution of Y can be used to form an interval estimate of c. Such an interval would be useful in deciding whether or not there are enough diseased individuals in the group screened to warrant screening again. The fact that the distribution of Y is in the same negative binomial family as X means that the same type of analysis could be applied to subsequent periods of screening. The conditional distribution of Y can also be used to construct probability statements on quantities calculated from c such as the underreferral proportion $c/[n - (a + b)]$.

When too little information is available about X to specify its mean and variance, α and β may be set equal to zero in the calculation of the parameters for the conditional negative binomial distribution of Y. This choice of parameters to represent a lack of information about X for this model is discussed by Hunter and Griffiths (2) in the context of the estimation of population size for insects. The methodology described here may also be applied to removal sampling over several periods in the animal sampling context described by Hayne (3), Moran (4), and Zippin (5) conditional upon p.

3. EXAMPLES

For the example from Grant introduced earlier, $E(X)$ would be taken to be 27 in accord with the national prevalence rate. The lack of information specific to the population being screened would be reflected by taking $V(X)$ large, say, equal to 3 times its mean. This corresponds to specifications of 13.5 and 0.5 for α and β, respectively. To facilitate comparisons with Grant's results, take $p = 0.74$, which is his estimate of the sensitivity of the screening procedure, although ideally p would be determined independently of the data on the program under evaluation. After observing that $a = 20$, the distribution of Y conditional upon a would have mean and variance 7.0 and 8.49, respectively. The mean of 7 agrees with Grant's estimate because we chose p to coincide with his data. However, a 95% probability interval for Y with a variance of 8.49 would extend from 3 through 14 inclusively. The corresponding interval for the underreferral rate would extend from 0.007 through 0.003. If $V(X)$ had been taken even larger, say, either 10 or 28 times $E(X)$, these intervals would still have had nearly as high a probability of holding, namely, 0.93. Intervals like these add greatly to an assessment of the screening program since they show that the underreferral rate could easily extend to twice its point estimate of 0.017. They were calculated utilizing tables of the binomial distribution (6) and the relationship between the positive and negative binomial distribution as described, for instance, in Raiffa and Schlaifer (7).

As another example, consider the evaluation of a screening program for breast cancer for asymptomatic women. For purposes of illustration, suppose that 1000 women were to be screened. Suppose experience with similar groups of women in other breast cancer detection centers indicates that the rate of breast cancer will be between 0.002 and 0.010. A reasonable choice for α and β is $\alpha = 20$ and $\beta = 4$, which gives a distribution with mean 0.0050 and standard deviation 0.0025 for the breast cancer rate and which leads to a probability of 0.916 for the rate (X/n) in the interval between 0.002 and 0.010. Suppose also that the screening procedure will be a combination of mammography and clinical examination and that screen positives will be followed with a biopsy. Assume that the sensitivity of the entire sequence of procedures if 0.9 $(=p)$. Now suppose that all 1000 women are screened in this manner and that 7 $(=a)$ women are found to have breast cancer. This leads to posterior parameters for the negative binomial distribution of $a + \alpha = 7 + 20 = 27$ and $(p + \beta)/(1 - p) = (0.9 + 4)/0.1 = 49$. The conditional mean of Y, which could be taken as a point estimate, is 0.55. Its conditional variance is 0.56, which leads to standard deviation of 0.75. Since the expected number of false negatives (Y) is less than one, the posterior probability of one or more missed cases capable of detection, that is, the $P(Y \geq 1)$, is of interest. The probability is 0.420. Also $P(Y \geq 2) = 0.107$ and $P(Y \geq 3) = 0.020$. These probabilities are quite sensitive to the choice of p. For instance, if p had been chosen to be 0.8 instead of 0.9, the corresponding probabilities would have been 0.625, 0.264, and 0.084, respectively. Although the negative binomial distribution was used in these evaluations, the Poisson distribution could have been used since the conditional variance of Y is only slightly larger than the mean and the negative binomial distribution approaches the Poisson as its variance decreases to its mean.

The probabilities calculated for this example could be easily converted into probability statements on the underreferral rate by dividing the limits on Y by the number of negative screens. More important, the probabilities of at least one, two, and three missed cases could be used, with information on cost-benefit ratios, to determine whether or not to rescreen the women who were negative on the first screen.

APPENDIX: CONDITIONAL DISTRIBUTION OF Y

Conditional upon the observed a, the probability function for X is

$$h(x|a) \propto f(a|x, p) g(x; \alpha, \beta)$$
$$\propto \Gamma(x + \alpha) (1 - p)^{(x-a)}[1/(1 + \beta)]^x/(x - a)!, \qquad x = a, a + 1, \ldots, \quad [2]$$

where $f(a|x, p)$ is binomial and $g(x; \alpha, \beta)$ is given by [1]. Since Y is the number of subjects remaining after removing a patients for treatment who were found to be positive by both the screening and diagnostic tests, $Y = X - a$ once a is observed. It follows from [2] that the probability function for Y conditional upon a is

$$w(y|a) \propto \Gamma(y + a + \alpha) \, [(1 - p)/(1 + \beta)]^y/\Gamma(y + 1), \qquad y = 0, 1, \ldots \quad [3]$$

Comparison of [3] with [1] shows that the distribution of Y conditional upon a is negative binomial with parameters α and β [1] replaced by $\alpha^* = a + \alpha$ and $\beta^* = (p + \beta)/(1 - p)$, respectively.

ACKNOWLEDGMENTS

REFERENCES

1. GRANT, J. A. Quantitative evaluation of a screening program. *Amer. J. Pub. Health* **64**, 66 (1974).
2. HUNTER, A. J., AND GRIFFITHS, H. J. Bayesian approach to estimation of insect population size. *Technometrics* **20**, 231 (1978).
3. HAYNE, D. W. Two methods for estimating populations from trapping records. *Journal of Mammal.* **30**, 399 (1949).
4. MORAN, P. A. P. A mathematical theory of animal trapping. *Biometrika* **38**, 307 (1951).
5. ZIPPIN, C. An evaluation of the removal method of estimating animal populations. *Biometrics* **12**, 163 (1956).
6. "Tables of the Binomial Probability Distribution." U.S. Department of Commerce, National Bureau of Standards, Washington, D.C., 1950.
7. RAIFFA, H., AND SCHLAIFER, R. "Applied Statistical Decision Theory," p. 236. Harvard Univ. Press, Boston, 1961.