

Estimating the sensitivity of a screening test

NICHOLAS E DAY

From the Unit of Biostatistics and Field Studies, International Agency for Research on Cancer, 69372 Lyon Cedex 08, France

SUMMARY A commonly used estimate of a screening test's sensitivity, based on the prevalence at screening and the incidence of interval cancers, is shown to be logically unsound and quantitatively poor. An alternative simple estimate is proposed, based only on incidence rates.

The purpose of this note is to draw attention to an estimate of the sensitivity of a screening test for the early detection of disease which is in common use and appears to be incorrect. The usual definition of the sensitivity of a screening test is in terms of the following schematic two by two table:

		True Disease State	
		+	–
Screening test	+	a	b
	–	c	d

The sensitivity is defined as the proportion of individuals with disease whom the screening test labels positive, that is,

$$\text{sensitivity} = S = a/(a+c)$$

The specificity is given by $d/(b+d)$. Estimation of specificity is a separate problem and for ease of presentation it will be assumed that the specificity is 100%.

Unless all individuals are subjected to a definitive diagnostic test, which in most situations would not be acceptable, the sensitivity is not directly estimable. In particular, the quantity c , the number of individuals screened negative who in fact have the disease, will not be known.

An estimate of sensitivity which is being increasingly used is based on the assumption that those incorrectly screened negative will surface with clinical manifestation of the disease in a short interval after the screen. An interval often used is 1 year.¹ The quantity c is then estimated by the number of cases of disease occurring in the year after screening, which will be denoted by c^* , and the estimate of sensitivity, \hat{S} say, given by

$$\hat{S} = \text{Estimate of sensitivity} = \frac{a}{a+c^*} \quad (1)$$

This estimate, which on first view might appear a reasonable approximation, can be seriously wrong. The

2×2 table given above, although attractively simple, ignores the time aspect of early detection. The lesions potentially identifiable by the screening test are those in the preclinical detectable phase, or PCDP. At a particular point in time the number of individuals in the PCDP is given by the incidence of the disease (I) multiplied by the average duration (μ) of the PCDP (assuming 100% specificity). The number of lesions detected will be the number that are prevalent multiplied by the sensitivity. That is, assuming $b=0$,

$$\text{Lesions detected at (first) screening} = S.I.\mu = a$$

In the year after screening, incident cases of disease will consist of those cases in the PCDP at screening but missed, and those cases whose PCDP began after the screening test, with duration less than a year. The estimate given by expression (1) has two sources of error. The denominator omits all cases which were missed at the screening test but do not surface until more than a year has elapsed, and includes cases which surface in the first year but were not in the PCDP at the time of the test. An assessment of the quantitative effect of these two sources of error can be made as follows.

The number of cases surfacing in the year after screening, whose PCDP began after the screening test, can be evaluated in terms of the distribution of the PCDP.² If $F(t)$ denotes the cumulative distribution function of the PCDP, thus giving the proportion of lesions with a PCDP less than or equal to t , the incidence (ie, integrated incidence rate) of new cases arising in the first year after screening is given by

$$I \int_0^1 F(t) dt$$

Of those surfacing at time t after the screening test who were missed at the test, the PCDP must have been greater than or equal to t . The incidence rate of these cases is therefore $I(1-S) [1-F(t)]$, and the overall incidence in the first year is given by

Estimating the sensitivity of a screening test

$$I(1-S) \int_0^1 (1-F(t))dt$$

The sum of these two components is thus

$$I(1-S) + IS \int_0^1 F(t)dt. \tag{2}$$

Substituting in expression (1) for \hat{S} gives

$$\begin{aligned} \hat{S} &= \frac{S.I.\mu}{S.I.\mu + I(1-S) + IS \int_0^1 F(t)dt} \\ &= S\mu / [S\mu + (1-S) + S \int_0^1 F(t)dt] \end{aligned} \tag{3}$$

The right-hand side is clearly not in general equal to S . How badly it performs as an estimate of S will depend mainly on the mean value μ of the PCDP but also in part on the form of $F(t)$. Both for breast cancer and for cancer of the cervix an exponential distribution fits reasonably well.^{3,4} Table 1 evaluates the right-hand side of expression (3) for various values of S and μ , assuming that $F(t)$ is exponential. It is clear that \hat{S} performs badly. Assuming alternative forms for $F(t)$ changes the entries in table 1 slightly, but not the overall picture.

Table 1 *Expected values of the estimate \hat{S} of sensitivity, under a range of conditions.*

True sensitivity	Average duration of the PCDP (years)			
	1	3	5	10
0.95	0.708	0.935	0.972	0.990
0.85	0.648	0.902	0.949	0.977
0.75	0.588	0.861	0.921	0.962

One can see from expression (2) that the sensitivity and the left-hand tail of the distribution of the PCDP both contribute to the incidence of early interval cancers. Methods have been proposed for obtaining separate estimates of S and $F(t)$,³ but the computation is laborious.

One approach to forming a simple estimate of S which performs better than \hat{S} might be to correct \hat{S} according to expression (3) to obtain a refined estimate \hat{S}_1 given by

$$\hat{S}_1 = \hat{S} / [\mu(1-\hat{S}) + \hat{S}(1-\int_0^1 F(t)dt)].$$

Unfortunately, this estimate requires knowledge both of μ and of $\int_0^1 F(t)dt$, and will be seriously in error if a wrong assumption is made for μ .

An alternative, which seems preferable, is to base a simple estimate of S on expression (2), the incidence of interval cases soon after the first screen. Denoting by I_T the incidence (ie, integrated incidence rate) in time T after the first screen, one has

$$I_T = I(1-S) + IS \int_0^T F(t)dt.$$

If an independent estimate is available of I , the incidence the women attending screening would have had in the absence of screening, which would be the case in a controlled trial, then an estimate of S , \hat{S}_2 say, is given by

$$\hat{S}_2 = \frac{1-(I_T/I)}{1-\int_0^T F(t)dt}.$$

This estimate is preferable to \hat{S}_1 , since it only requires knowledge of $\int_0^T F(t)dt$, and not of μ as well.

An additional reason for preferring \hat{S}_2 to \hat{S} is that the latter is affected by lack of specificity of the screening test. Many of the lesions detected at screening may never proceed to clinical disease, so that the screening prevalence is overestimated. The estimate \hat{S}_2 refers strictly to those lesions present at screening which would progress to clinical disease.

$\int_0^T (t)dt$ is simply the proportion of cases with aPCDP of

duration less than time T . In the absence of any knowledge of $F(t)$, one can go no further than asserting $1-I_T/I \leq \hat{S}_2 \leq 1$. When screening for breast cancer, however, a reasonable approximation would be that F is exponential with mean about two years.⁵ For cervical cancer screening, F could be taken as exponential with mean 15 years.⁴ If one takes T to be fairly small compared to the estimate of the mean, then the denominator of \hat{S}_2 will be close to one, and relatively insensitive to errors in the estimate of the mean.

Assuming the distribution of the PCDP to be exponential with mean two years, and taking the incidence in the first six months after screening, then

$$\int_0^{0.5} F(t)dt = 0.0576$$

and

$$\hat{S}_2 = \frac{1-(I_T/I)}{0.9424}.$$

Table 2a shows the result of using this estimate for various values of S if the correct value of the mean differs from 2.

Table 2b gives corresponding results when an interval of one year is used, for which

$$\hat{S}_2 = \frac{1 - (I_T/I)}{0.7869}$$

Comparison of table 2a and 2b indicates that for false negative rates of 15% or less an interval of one year may give considerably biased estimates, whereas the interval of six months performs well.

Table 2 Expected values of the estimate \hat{S}_2 under a range of conditions (breast cancer screening)

(a) Using 6-month interval cases

Sensitivity	Mean value of the PCDP (in years)		
	1.5	2.0	2.5
0.95	0.933	0.95	0.961
0.85	0.835	0.85	0.860
0.75	0.736	0.75	0.759

$$\hat{S}_2 = (1 - (I_T/I))/0.9424$$

(b) Using 12 months interval cases

Sensitivity	Mean value of the PCDP (in years)		
	1.5	2.0	2.5
0.95	0.881	0.95	0.995
0.85	0.788	0.85	0.890
0.75	0.696	0.75	0.786

$$\hat{S}_2 = (1 - (I_T/I))/0.7869$$

For cancer of the cervix, assuming a 15 year mean PCDP and basing the estimate of sensitivity on one year of incidence, gives

$$\hat{S}_2 = (1 - (I_T/I))/0.9674.$$

Table 3 gives the value of \hat{S}_2 for various values of S and of the mean PCDP. The performance is reasonably good.

Table 3 Expected values of the estimate \hat{S}_2 under a range of conditions (cervical cancer screening)

Using 12 months interval cases

Sensitivity	Mean value of the PCDP (in years)		
	10	15	20
0.95	0.935	0.95	0.958
0.85	0.836	0.85	0.857
0.75	0.738	0.75	0.756

$$\hat{S}_2 = (1 - (I_T/I))/0.9674$$

These two proposed estimates of sensitivity should not be regarded as being in any sense optimum. They are proposed solely as simple estimates with a more secure logical foundation than the commonly used estimate \hat{S} .

References

¹Chamberlain J, Clifford RE, Nathan BE, Price JL, Burns I. Repeated screening for breast cancer. *J Epidemiol Community Health* 1984; **38**: 54-7.
²Day NE, Walter SD. Simplified models of screening for chronic disease from mass screening programmes. *Biometrics* 1984; **40**: 1-14.
³Walter SD, Day NE. Estimation of the duration of a preclinical disease state using screening data. *Am J Epidemiol* 1983; **118**: 865-85.
⁴Brookmeyer R, Day NE, Moss S. Case-control studies for the estimation of the natural history of preclinical disease. Submitted for publication.
⁵Day NE, Walter SD, Collette HJA. Statistical models of disease natural history: their use in the evaluation of screening programmes. In: Prorok PC, Miller AB, eds. *Screening for cancer. I - General principles on evaluation of screening for cancer and screening for lung, bladder and oral cancer*. Geneva: International Union Against Cancer (UICC Technical Report Series No. 78), 1984, 55-70.