

Quantifying the value of markers in screening programmes

Søren Dinesen Østergaard · Peter Thisted Dinesen ·
Leslie Foldager

Received: 16 November 2009 / Accepted: 20 January 2010 / Published online: 12 February 2010
© Springer Science+Business Media B.V. 2010

Abstract Existing methods used to rank the value of individual screening markers in screening programmes are inadequate. We have developed a simple Screening Marker Index: (Screening Marker Index = Positive Predictive Value \times Sensitivity). The Screening Marker Index proved to be superior to existing indices in ranking screening markers according to their ability to identify the conditions sought.

Keywords Screening · Sensitivity · Specificity

Abbreviations

COPD	Chronic obstructive pulmonary disease
NNS	Number needed to screen
PPV	Positive predictive value
SMI	Screening marker index
YI	Youden index
PLR	Positive likelihood ratio
NLR	Negative likelihood ratio
ROC	Receiver operating characteristics

S. D. Østergaard (✉)
Unit for Psychiatric Research, Aalborg Psychiatric Hospital,
Aarhus University Hospital, Mølleparkvej 10,
9000 Ålborg, Denmark
e-mail: soerendinesen@gmail.com

P. T. Dinesen
Department of Political Science, Aarhus University,
Århus, Denmark

L. Foldager
Centre for Psychiatric Research, Aarhus University Hospital,
Risskov, Denmark

L. Foldager
Bioinformatics Research Centre (BiRC), Aarhus University,
Århus, Denmark

Screening for diseases has become increasingly common at all levels of modern health care [1, 2]. The tendency to screen has increased due to advances in the quality and numbers of available treatments, combined with more accurate diagnostic tools [3].

Under the right circumstances screening can detect disease before it becomes severe and thereby decrease mortality and improve health [4–6]. However, screening should be used with caution to avoid unnecessary diagnostic testing, false positive diagnoses and waste of health care resources [2]. To ensure that screening initiatives are justified from ethical and pragmatic points of view the WHO has defined a set of criteria to be fulfilled prior to the initiation of a screening programme. These criteria are based on the theories of Wilson and Jungner [7]. A selection of the WHO-criteria are listed below [3]:

- The condition sought should be an important health problem.
- There should be an accepted treatment for patients with recognized disease.
- There should be a suitable test or examination.
- The test should be acceptable to the population.
- The cost of case-finding (including diagnosis and treatment of patients diagnosed) should be economically balanced in relation to possible expenditure on medical care as a whole.

As a consequence of the latter criteria, case-finding has to be efficient in order to minimize costs in the health care system. Efficiency in case-finding is often described by the “number needed to screen” (NNS) [8, 9]. The NNS is defined as the “the number of people that need to be screened for a given duration to prevent one death or adverse event” [10]. The adverse event can be interpreted differently according to the setting of the screening

Table 1 The confusion matrix

	Disease—actual value		
	Present (D)	Absent (D)	
Marker—predicted outcome			
Present (M)	a (True positives)	c (False positives)	a + c
Absent (M)	b (False negatives)	d (True negatives)	b + d
	a + b	c + d	N = a + b + c + d

programme. For instance the NNS can be defined as the average number of people that would need to be screened to detect one case of a particular disease.

In the planning of screening initiatives it is essential to assess which markers that are correlated with the condition sought [3]. The extent of this correlation is embodied in the NNS, which expresses the inverse of the positive predictive value (PPV) for a marker and hence should be as low as possible. However, a low NNS does not necessarily ensure that a marker has practical value in a screening programme. If the combined presence of the marker and the condition sought is very rare, a low NNS has no practical significance. Screening in the presence of such a marker would detect very few cases. In other words, two markers with identical NNS do not necessarily represent the same value in a screening programme. The exact same problem arises using the Youden index (YI) [11] and the positive-/negative likelihood ratios (PLR/NLR), other measures commonly used to describe the efficiency of diagnostic tools.

Another method to evaluate a screening marker is by making a graphical plot of the sensitivity versus 1-specificity in a coordinate system which defines the receiver operating characteristic (ROC). Like the indices mentioned above, the ROC is also related to cost/benefit analysis of diagnostic decision making. Since the ROC is also based on sensitivity/specificity it possesses the same weaknesses as the YI, the PLR and the NLR when it comes to the hierarchical ranking of screening markers. This is illustrated by the example given below.

In conclusion, existing indices seem inadequate in the comparison of individual screening markers.

In order to facilitate comparison of the value of various screening markers in potential/existing screening programmes we have developed a simple Screening-Marker Index (SMI). The Screening-Marker Index is defined as follows:

$$\text{SMI} = P(M|D) \times P(D|M) = \text{sensitivity} \times \text{PPV}$$

The elements in the equation above are explained by the following 2×2 table (Table 1), also known as the confusion matrix. It follows that:

$$P(D) = \frac{a+b}{N} = \text{prevalence}$$

$$P(M|D) = \frac{a}{a+b} = \text{sensitivity}$$

$$P(-M|-D) = \frac{d}{c+d} = \text{specificity}$$

$$P(D|M) = \frac{a}{a+c} = \text{positive predictive value (PPV)}$$

Here “-” denotes negation, in the sense that “-D” means: “disease absent”. The NNS is the inverse of the PPV, i.e. $\text{NNS} = 1/\text{PPV}$.

The YI represents the marker’s “success” in predicting disease among the diseased and vice versa for subjects where disease is absent:

$$\text{YI} = \frac{1}{2} \left\{ \left(\frac{a}{a+b} - \frac{b}{a+b} \right) + \left(\frac{d}{c+d} - \frac{c}{c+d} \right) \right\}$$

It is straight forward to show that this index can be calculated as:

$$\text{YI} = \text{sensitivity} + \text{specificity} - 1$$

We note that this index weighs sensitivity and specificity equally. Usually YI is used only under the assumption that the proportion of true positives (sensitivity) is greater than the proportion of false positives (1-specificity). If this is not the case, then YI becomes negative; and to cite Youden: “Such a test is obviously worthless” [11]. However, in the present setting, we are also interested in ranking markers with low sensitivity and specificity, with negative YI-values as a consequence. The range for the YI is therefore (-1 to 1) whereas the range for SMI, being a product of two probabilities, is (0 to 1). Note also, that the SMI (like YI) is one for the “ideal” marker where both b and c is zero.

Likelihood ratios indicate how much a particular test-result should shift the suspicion towards the presence or absence of disease. The ratios can also be used to describe the value of a screening marker. In that setting the PLR indicates how much the suspicion of disease should be increased if the screening marker is present. Conversely, the NLR indicates how much the suspicion of disease should be decreased if the marker is absent. The PLR and NLR are defined as follows:

$$PLR = \frac{\text{sensitivity}}{(1 - \text{specificity})}$$

$$NLR = \frac{1 - \text{sensitivity}}{\text{specificity}}$$

The PLR and the NLR range from zero to infinite. The ideal screening marker has a high PLR and a low NLR.

To visualize and compare the performance of each index we considered three fictional diseases with prevalences of 1, 10 and 20% in a population of 1,000 individuals. We then calculated the NNS, YI and SMI for all possible combinations of specificity and sensitivity between zero and one. The result is shown in the contour plots below (Fig. 1).

We note that the NNS favours markers with high sensitivity, the YI weights sensitivity and specificity equally whereas the SMI favours markers with high sensitivity without neglecting the importance of high specificity.

The following example demonstrates the practical differences between the markers;

The British National Institute of Clinical Excellence (NICE) has proposed screening for depression in primary

care and general hospital settings in high-risk groups [12]. Both chronic obstructive pulmonary disease (COPD) and heart disease are associated with the presence of depression [13, 14]. The Danish National Board of Health has concretized the NICE recommendations proposing COPD and heart disease as screening markers for depression in the primary care setting (along with other diseases) [15]. We now wish to determine whether COPD or heart disease is the better screening marker for depression.

Let's consider 1,000 (fictional) individuals, of which 200 suffer from depression, 100 has heart disease, 20 has both heart disease and depression, 250 have COPD and 50 have both COPD and depression. It follows that:

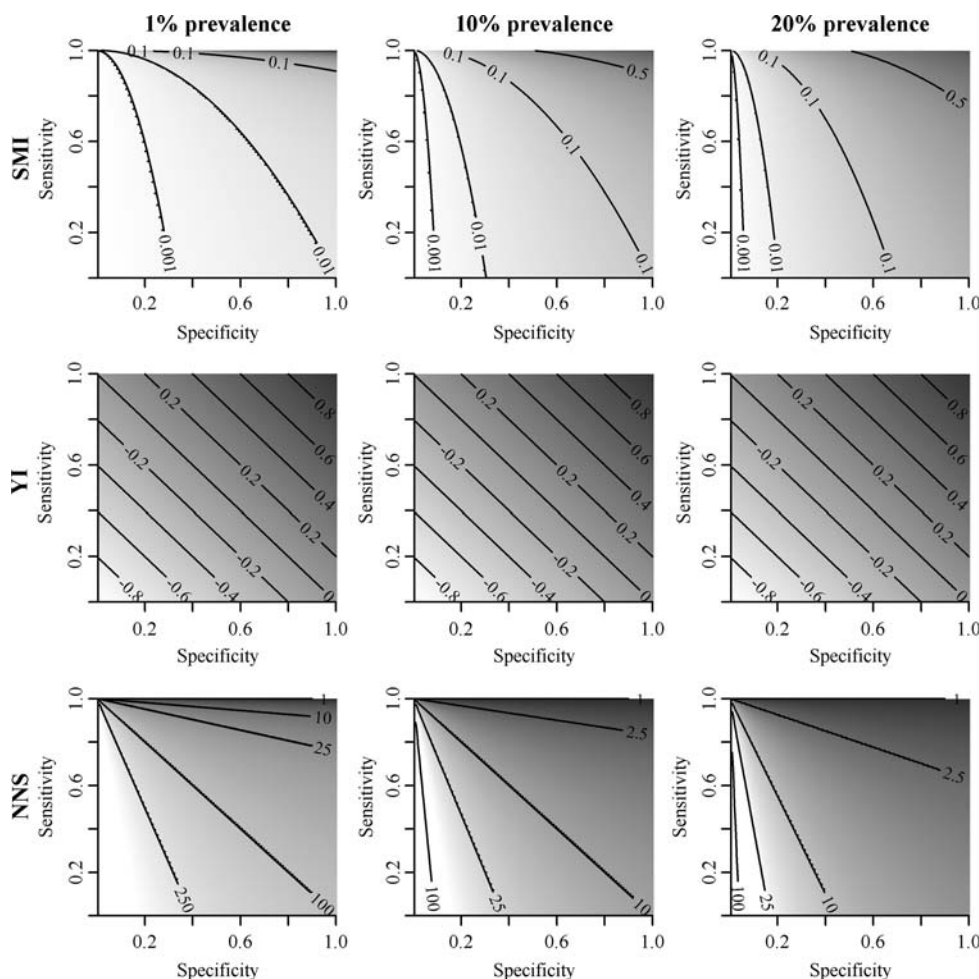
Heart Disease:

$$NNS = 100/20 = 5 \quad YI = 0.10 + 0.90 - 1 = 0$$

$$SMI = 0.10 \times 0.2 = 0.02$$

$$PLR = 0.10/(1 - 0.90) = 1 \quad NLR = (1 - 0.10)/0.9 = 1$$

Fig. 1 Contour plots showing the values of NNS, YI and SMI for screening markers using combinations of sensitivity and specificity between 0 and 1. The three indices are displayed for three diseases with prevalence of 1% (left), 10% (middle) and 20% (right). Grayscale colours have been added to facilitate visual interpretation with the scale ranging from white (bad marker) to dark gray (good marker)



COPD:

$$\text{NNS} = 250/50 = 5 \quad \text{YI} = 0.25 + 0.75 - 1 = 0$$

$$\text{SMI} = 0.25 \times 0.2 = 0.05$$

$$\text{PLR} = 0.25/(1 - 0.75) = 1 \quad \text{NLR} = (1 - 0.25)/0.75 = 1$$

The ROC for the two examples are not depicted, but the coordinates have the same distance to the straight line through (0,0 and 1,1) indicating equal value as diagnostic marker.

This demonstrates how the existing indices are insensitive to the fact, that screening for depression among people with COPD would detect 2.5 as many cases of depression, than screening in the presence of heart disease. The SMI is sensitive to this difference.

We have shown that existing indices used to compare the value of individual screening markers in screening programmes possess some inherent weaknesses. Our simple SMI circumvent these problems to a large extent.

As demonstrated by the contour plots in Fig. 1, identical SMI values can occur with very different combinations of sensitivity and specificity. However, as outlined in the example on depression-screening, this problem has no consequences in clinically relevant situations.

In practical terms the SMI can be used to select which markers that should elicit screening and thereby contribute to sensitive diagnosis of various diseases. This is in accordance with the WHO guidelines on screening [3].

References

1. Love RR, Camilli AE. The value of screening. *Cancer*. 1981;48:489–94.
2. Grimes DA, Schulz KF. Uses and abuses of screening tests. *Lancet*. 2002;359:881–4.
3. Andermann A, Blancquaert I, Beauchamp S, Dery V. Revisiting Wilson and Jungner in the genomic age: a review of screening criteria over the past 40 years. *Bull World Health Organ*. 2008;86:317–9.
4. Sasieni P, Castanon A, Cuzick J. Screening and adenocarcinoma of the cervix. *Int J Cancer*. 2009;125:525–9.
5. Hofvind S, Geller B, Vacek PM, Thoresen S, Skaane P. Using the European guidelines to evaluate the Norwegian Breast Cancer Screening Program. *Eur J Epidemiol*. 2007;22:447–55.
6. Irwig L, Macaskill P, Houssami N. Evidence relevant to the investigation of breast symptoms: the triple test. *Breast*. 2002;11:215–20.
7. Wilson JM, Jungner YG. Principles and practice of mass screening for disease. *Public Health Papers* 34. Geneva: World Health Association; 1968.
8. Cuijpers P, van Straten A, Smits N, Smit F. Screening and early psychological intervention for depression in schools: systematic review and meta-analysis. *Eur Child Adolesc Psychiatry*. 2006;15:300–7.
9. Storm-Versloot MN, Vermeulen H, Wiggers LC, Smets EM, de Haes HC, Peters RJ, et al. The number of smokers needed to screen and treat in a smoking cessation programme. *Eur J Cardiovasc Prev Rehabil*. 2009;16:669–76.
10. Rembold CM. Number needed to screen: development of a statistic for disease screening. *BMJ*. 1998;317:307–12.
11. Youden WJ. Index for rating diagnostic tests. *Cancer*. 1950;3:32–5.
12. National Institute of Clinical Excellence. Clinical Guideline 23 (amended), Depression: management of depression in primary and secondary care—NICE guidance. London 2007. Available at: <http://guidance.nice.org.uk/CG23/Guidance/pdf/English>.
13. Sorensen C, Brandes A, Hendricks O, Thrane J, Friis-Hasche E, Haghfelt T, et al. Depression assessed over 1-year survival in patients with myocardial infarction. *Acta Psychiatr Scand*. 2006;113:290–7.
14. Stage KB, Middelboe T, Stage TB, Sorensen CH. Depression in COPD—management and quality of life considerations. *Int J Chron Obstruct Pulmon Dis*. 2006;1:315–20.
15. The Danish National Board of Health. Referenceprogramme on the management of unipolar depression in adults (in Danish). Secretariat for Referenceprogrammes 2007. Available at: http://www.sst.dk/publ/Publ2007/PLAN/SfR/SST_Dep. rapport.pdf.