Wolters Kluwer Health | Lippincott Williams & Wilkins

EUROPEAN SOCIETY OF CARDIOLOGY®

Review

Improvement of cardiovascular risk prediction: time to review current knowledge, debates, and fundamentals on how to assess test characteristics

Michel Romanens[a], Franz Ackermann[b], John David Spence[c], Roger Darioli[d], Nicolas Rodondi[d], Roberto Corti[e], Georg Noll[e], Matthias Schwenkglenks[f] and Michael Pencina[g]

[a]Cantonal Hospital, [b]Vascular Risk Foundation, Olten, Switzerland, [c]Stroke Prevention and Atherosclerosis Research Centre, London, Ontario, Canada, [d]Department of Ambulatory Care and Community Medicine, University of Lausanne, Medical Outpatient Clinic, University Hospital, Lausanne, [e]Cardiology Department, University Hospital, Zurich, [f]Head European Center of Pharmaceutical Medicine (ECPM) Research, Basel, Switzerland and [g]Department of Mathematics and Statistics, Framingham Heart Study, Boston University, Boston, Massachusetts, USA

Received 7 April 2009 **Accepted** 21 October 2009

Cardiovascular risk assessment might be improved with the addition of emerging, new tests derived from atherosclerosis imaging, laboratory tests or functional tests. This article reviews relative risk, odds ratios, receiver-operating curves, posttest risk calculations based on likelihood ratios, the net reclassification improvement and integrated discrimination. This serves to determine whether a new test has an added clinical value on top of conventional risk testing and how this can be verified statistically. Two clinically meaningful examples serve to illustrate novel approaches. This work serves as a review and basic work for the development of new guidelines on cardiovascular risk prediction, taking into account emerging tests, to be proposed by members of the 'Taskforce on Vascular Risk Prediction' under the auspices of the Working Group 'Swiss Atherosclerosis' of the Swiss Society of Cardiology in the future. ***Eur J Cardiovasc Prev Rehabil*** **00:000–000 © 2010 The European Society of Cardiology**

European Journal of Cardiovascular Prevention and Rehabilitation 2010, **00**:000–000

Keywords: myocardial infarction, risk prediction, statistical methods

Introduction

In view of the many new possibilities to estimate cardiovascular risk in primary care, the executive committee of the working group on lipids and atherosclerosis of the Swiss Society of Cardiology decided to endorse a new taskforce on vascular risk prediction. During a creative process that started in February 2005, the taskforce reviewed the evidence from the literature that reflects current knowledge.

For physicians working in primary care, prevention of cardiovascular diseases in their patients has been an important goal. This is reflected in the new global risk assessment algorithms issued by the International Atherosclerosis Society (IAS [1,2]) and European Atherosclerosis Society (EAS [3]), which serve as guidelines. They are based on data from large populations, are statistically valid, and seem to be especially helpful to identify low-risk patients because of a very high specificity (SP) (> 90% [2]). Doubt remains about who should receive intensive primary prevention therapies in primary care, as global risk assessment tools have a rather low sensitivity (SE) of approximately 35% to detect patients at a high risk for cardiovascular events in the future, when currently defined cut-offs for high risk are used [2].

Risk stratification tools such as tables and risk calculators (see *www.agla.ch* adapted for the Swiss population) are

Correspondence to Michel Romanens, MD, Consultant in Cardiology, Ziegelfeldstr. 1, CH-4600 Olten, Switzerland
Tel: +41 62 212 44 10; fax: +41 62 212 44 30; e-mail: info@kardiolab.ch

1741-8267 © 2010 The European Society of Cardiology DOI: 10.1097/HJR.0b013e3283347059

based on data from large populations and usually provide valid statistical algorithms for these populations (internal validity), but before generalizing to new populations, their external validity should be determined. Although these tools are essential for an easy stratification, room for improvement does exist. Imaging, laboratory, and functional tests might have the potential to add information to improve cardiovascular risk assessment.

We aim at providing a comprehensive overview of current issues associated with the integration of emerging risk prediction tools. Our aim is to answer the question of what statistical instruments can be used to judge if a new test (e.g. atherosclerosis imaging of carotid arteries) increases the ability to predict risk where the results of conventional risk factor-based testing are already known. In other words, how do we assess if an additional test has an incremental value?

In the subsequent sections, we revisit key definitions of test characteristics; review different statistical methods to assess the clinical benefit of an additional test and briefly address the problem of who should be administered an additional test (the entire population potentially eligible for primary prophylaxis or more limited subpopulations, individuals only?); and touch upon economic considerations.

Key definitions of test characteristics

In situations where there exists a universally accepted risk threshold, the performance of a new test may be quantified in relatively simple terms. A test result can be positive, indicating the presence of disease, or it can be negative, indicating the absence of disease. As a test may give erroneous results, it may be true positive (TP), false positive (FP), true negative (TN), and false negative (FN). This is the basic concept of the two-by-two table in its application to diagnostic tools.

Further, knowing these key variables for a given test allows expressing its performance in terms of SE, SP, positive predictive value (PPV) and negative predictive value (NPV), positive (pLR) and negative likelihood ratio (nLR) and accuracy (ACC) [[4–7], (Table 1)]. SE is defined by the rate of TP in a positive test result [TP/(TP + FN)]. SP is defined by the rate of TN in a negative test result [TN/(TN + FP)]. Although SE and SP are independent of the prevalence of a disease, PPV and NPV are dependent on the pretest probability, which in multivariable models would be identical to the predicted incidence of a disease for a given configuration of risk factors. PPV is the rate of TP in all positive test results [TP/(TP + FP)], whereas NPV is the rate of TN in all negative test results [TN/(TN + FN)]. ACC is defined as TP + TN divided by the total population size and is also dependent on the prevalence of a disease.

Table 1 Calculation of test performances (modified from Ref. [7])

SE = TP/(TP + FN)
SP = TN/(TN + FP)
PPV = TP/(TP + FP)
NPV = TN/(TN + FN)
pLR = SE/(1 – SP)
nLR = (1 – SE)/SP
ACC = (TP + TN)/(TP + TN + FP + FN)

For the calculation of probabilities, a range from zero to 1.00 is used. Multiply by 100 to obtain percentages and accordingly, sensitivities and specificities are not expressed in percent but in percent divided by 100. ACC, accuracy; FN, false negative; FP, false positive; nLR, negative likelihood ratio; NPV, negative predictive value; pLR, positive likelihood ratio; PPV, positive predictive value; SE, sensitivity; SP, specificity; TN, true negative; TP, true positive.

The preference for reporting SE and SP versus PPV and NPV as primary summary characteristics varies in the scientific community. If we want to know what fraction of true events we will discover with a given test, then SE/SP is preferred. However, if we reverse the question and ask, given that my patient is test-positive, what are her chances of developing the condition, PPV and NPV can answer this directly. Summary measures, like for example, the Youden index (SE + SP–1), which serve as the basis of integrated discrimination improvement (IDI), are yet another option.

A good estimate for the performance of a test would be given by a change of the pretest probability using likelihood ratios. Likelihood ratios can be calculated by using the SE and SP of a test: pLR [SE/(100–SP)] and nLR [(100–SE)/SP]. As an example, a test with an SE of 30% and an SP of 90% (which approximately corresponds to the performance of the PROCAM or the SCORE algorithm) yields a pLR of [30/(100–90) = 3.00] and an nLR of [(100–30)/90 = 0.78]. As we can see now, global cardiovascular risk estimates have an acceptable pLR, but a rather weak nLR, despite their reportedly high NPV of about 90%, which is due to the relatively low overall probability for heart attacks in the population.

How to assess the quality of an additional test?

One main problem inherent to cardiovascular risk prediction stems from the need to integrate test results into medical therapeutic decisions. This problem is amplified in preventive medicine as the paradigm of classical medicine – 'establishing a diagnosis' – is replaced by 'estimating a risk for a diagnosis' that may or may not occur in the future. A national or international consensus defines the cut-off level for a high-risk situation. Although high-risk patients do not have a 'disease' *per se*, they often receive intensive and costly medical therapies. In Switzerland, people having diabetes or an IAS risk of greater than 20% in 10 years have a high risk for vascular events, for example, myocardial infarction [1]. Therefore, they are treated intensively to lower their risk.

As a first step in the primary cardiovascular prevention setting, risk is always assessed using coronary risk charts provided by IAS [1] or by EAS [3]. However, as previously mentioned, both the IAS and the EAS algorithm have a relatively low SE (30%) to predict the diagnoses of 'myocardial infarction' and 'vascular death', respectively, within the next 10 years, when the defined cut-offs for high risk are used [2,3]. Therefore, approximately 70% of patients who will eventually develop a vascular event are being missed using conventional cardiovascular risk factor algorithms. The hope is that emerging risk factors might help to close this detection gap, at least in part [8,9], with the help of laboratory tests (e.g. high-sensitivity C-reactive protein, homocysteine, plasminogen matrix metalloproteinase-1, inter-cellular adhesion molecule 1), arterial function tests (e.g. endothelial function tests, ankle-brachial-index, tests for aortic stiffness or elastance), and imaging tests of atherosclerosis (carotid intima-to-media thickness, carotid plaques, coronary calcifications).

There are various ways in which results from these emerging cardiovascular risk factors might be integrated into clinical practice.

The question is, do they add significantly to risk prediction? We can distinguish between methods to evaluate model performance, specifically c-statistics, IDI, net reclassification improvement (NRI), and classification rules (posttest odds):

(1) Measures of relative risk: relative risk, odds ratios, hazard ratios.
(2) Measures of model performance: c-statistics, calibration χ^2.
(3) Measures of model improvement: difference in c-statistics, NRI, IDI.
(4) Methods to add new test: posttest risk calculation, redevelopment of the model with new test variable.

Risk ratios, odds ratios and hazard ratios calculate the magnitude of effect between different values or categories of the new test, and when used in multivariable models, help to determine independent association of the new test with outcome. Obtaining precise estimates of these measures of relative risk is one of the necessary conditions for good discrimination of the model. Discrimination, often quantified using the c-statistics, measures the model's ability to distinguish cases from non-cases. A complementary measure, called calibration, assesses the model's ability to assign risks that are close to those observed in practice. Discrimination can be more tied to relative risk, while calibration to absolute risk, but these two are usually closely related. C-statistics calculate the accuracy of the discrimation between subjects who will and who will not develop a disease [10]. An integration of test accuracy expressed as calibration and discrimination is the posttest risk calculated with likelihood statistics, where the probability of developing a disease is integrated by using the pretest probability. Both likelihood statistics and newly developed reclassification systems for risk categories (NRI) and for continuously described risk (IDI) allow the shift of a subject from one risk category to another, such as from low risk to intermediate risk. For example, using NRI, the accuracy of a test is measured by the number of subjects in whom a correct risk category change occurs. If such a test does not correctly change risk categories in a significant proportion of subjects, this test does not appear to be clinically meaningful.

Table 2 Criteria for a good screening test (modified from Ref. [11])

1. Independent comparison with a gold standard
2. Large spectrum of pretest probabilities
3. Ability to change clinical decisions
4. High reproducibility
5. Validation in several populations
6. High accuracy to discriminate individuals with and without disease

The basic requirements that a new test should satisfy before being considered for inclusion in clinical practice are given in Table *2* [11].

Relative risk indicators

Risk ratios, odds ratios or hazard ratios are frequently used in the literature to assess the performance of a test. Hazard ratios, for example, for high-sensitivity C-reactive protein or intima-to-media thickness result frequently in statistically significant improvements in risk prediction [12,13] and are based statistically on a multiple logistic regression model. However, when receiver-operating characteristic (ROC) analysis is applied to the same tests, they usually do not show a statistically significant improvement [10,14–16]. This problem is still open to debate and a final answer to this goes beyond our review. However, as suggested by Cook and Pencina [10,17], one solution is the calculation of posterior risk and to look at the number of reclassified subjects.

Receiver-operating characteristic analysis based on c-statistics

In the binary case, ROC curves (c-statistic) display a plot of SE versus 1 – SP over all possible risk thresholds after inclusion of a continuous predictor. Area under the ROC curve (AUC) has been used as a measure of discrimination, namely the model's ability to distinguish future events from non-events based on baseline risk prediction. A test with an AUC of 0.70 or more may be useful. The incremental value of a second test is determined by the combined AUC of both tests (e.g. 0.82), which would in this example correspond to an additional diagnostic information of 0.12 (or 12%). This is an example of a clinically meaningful increase in the AUC when a new

test is added. The difference may become statistically significant (depending on the sample size, the number of events observed [18]). ROC analysis has however been criticized for being a too conservative statistical approach and therefore rejects many new emerging tests for the prediction of cardiovascular events despite clinically important and statistically significant increased relative risks after adjustment for a variety of other cardiovascular risk factors [10].

Beyond ROC: net reclassification improvement and integrated discrimination improvement

A promising step ahead would be to calculate the ability of a new emerging test to reclassify subjects to a different risk category. The ability of a risk marker to accurately stratify individuals into higher or lower risk categories (reclassification) is increasingly used in the literature on behalf of the NRI (categorical) and IDI (continuous) [17].

The following example is based on Ref. [17]. Consider a model for coronary heart disease (CHD) risk prediction which contains standard risk factors of the Framingham Risk Score (age, male sex, systolic blood pressure, smoking, diabetes and total cholesterol). The question is whether the addition of HDL cholesterol improves prediction in a meaningful way. Among 3264 individuals free of cardiovascular disease at baseline, 183 develop CHD during 10 years of follow-up. On the basis of the National Cholesterol Education Program ATPIII guidelines, people with CHD risk above 20% would be targeted for aggressive treatment and those below 6% would be considered low risk. Everyone can be classified into one of the three risk categories based on a pretest probability derived from a model without HDL. If a new test (HDL in this case) is useful, models which additionally contain HDL should move those who will experience events upwards to a higher risk category and those who will not, downwards. The reclassification for this example is presented in Table 3. We see that 29 events were moved up and 7 down for a net gain of 22 out of a total of 183. Among 3081 nonevents, 174 were moved down and 173 up for a net gain of 1. The NRI is given as $22/183 + 1/3081 = 12.1\%$. Methods exist to construct a confidence interval and perform a test of significance for the net effect.

The IDI is a continuous version of the net reclassification. Here, instead of counting each movement in the right direction as one and in the wrong direction as minus one, we calculate the difference in predicted probabilities between models with and without the new test and add them separately for events and non-events.

Example of posttest risk odds using total plaque area of carotid arteries

Posttest risk odds are helpful to reclassify a patient risk, for example, from intermediate to high. The posterior probability of disease is calculated based on the pretest odds multiplied by the likelihood ratio (for details refer to Table 4), and is a conservative but widely accepted estimate of posttest risk [19–21]. This kind of risk prediction is, however, based on the assumption that the pretest probability of risk can be determined with sufficient precision and may not adequately account for correlations between the standard risk factors and the new test [22].

In the setting of cardiovascular primary prevention, the posttest risk approach allows the use of a second or sequential test to calculate posttest probabilities (PTPs), whereas the pretest or first test is based, for example, on the IAS algorithm. In the following example we used the IAS risk of individuals and total plaque area (TPA) to calculate a posttest risk. TPA is a measure of the total plaque burden of the carotid arteries. Plaques are traced longitudinally,

Table 3 Reclassification among people who experience a CHD event and those who do not experience a CHD event on follow-up

Model without HDL	Model with HDL			Total
Participants who experience a CHD event				
Frequency row percentage	<6%	6–20%	>20%	
<6%	39	15	0	54
	72.22	27.78	0.00	
6–20%	4	87	14	105
	3.81	82.86	13.33	
>20%	0	3	21	24
	0.00	12.50	87.50	
Total	43	105	35	183
Participants who do not experience a CHD event				
Frequency row percentage	<6%	6–20%	>20%	
<6%	1959	142	0	2101
	93.24	6.76	0.00	
6–20%	148	703	31	882
	16.78	79.71	3.51	
>20%	1	25	72	98
	1.02	25.51	73.47	
Total	2108	870	103	3081

In Table 3, among those experiencing a coronary event, 54 patients were classified as low risk. Adding HDL as a new test, 15 patients were shifted into the intermediate risk category. In the intermediate risk group (e.g. 6–20%), four were incorrectly moved to low risk by HDL, but 14 patients were correctly shifted into the high-risk group (net gain: 10 patients). All over, $15 + 10 - 3 = 22$ patients were correctly reclassified by HDL. In analogy, in Table 3, only one additional patient was correctly reclassified by HDL. In this example, the net reclassification is given as $22/183 + 1/3081 = 12.02 + 0.03\% = 12.1\%$.

Table 4 Formula for the calculation of posttest probabilities

PTP pos: $(PV \times SE)/[PV \times SE + (1 - PV) \times (1 - SP)]$
PTP neg: $[PV \times (1 - SE)]/[PV \times (1 - SE) + SP \times (1 - PV)]$

This formula is essentially an abbreviation for the calculation of posttest risk based on the pretest odds multiplied by the likelihood ratios. As an example, pretest probability may be 0.33 or 33%, therefore pretest odds are $0.33/1 - 0.33 = 1/2$. Likelihood ratios are given in Table 5 and are based on the sensitivity and specificity of a test for different cut-offs. As an example, patients with carotid atherosclerosis within the fourth quartile have a positive likelihood ratio of 2.52. Posttest odds are therefore: $2.52/2 = 1.26$. Posttest odds are converted into a posttest probability by $1.26/1.26 + 1 = 0.56$. In summary, risk of 33% increases to 56% in this example. PTP neg, posttest probability for a disease if the test is negative (normal); PTP pos, posttest probability for a disease if the test is positive (pathologic); PV, pretest probability [or prevalence (PV)] for a disease; SE, sensitivity; SP, specificity.

and the TPA is derived from the sum of all plaque areas detected during the imaging of both carotid arteries. Imaging is performed with the patients in the supine position as described in the original Canadian London cohort [23]. Within this cohort, we identified 684 originally healthy individuals, in whom 13 fatal and nonfatal myocardial infarctions were observed during a follow-up period of 3 years (unpublished data). Furthermore, we determined the SE and SP of the quartiles of TPA to detect these 13 myocardial infarctions (Table 5). On the basis of the posttest risk formula (Table 4), posttest risk can be calculated. This approach is exemplified in Table 6, where an IAS risk for myocardial infarction in 10 years is taken as the pretest probability and TPA is used to calculate the TPA-PTP. In addition, 95% confidence intervals are calculated for this posttest risk, for different levels of pretest risk (as an example: 5, 10, and 15% pretest probabilities based on IAS). The results in Table 6 show that this kind of sequential testing is useful to identify high-risk patients with intermediate IAS risk; for example, 10–19% 10-year risk for myocardial infarction. If in a given patient, the IAS risk was 15% and TPA was above 0.18 cm^2, then this patient would be reclassified from the intermediate to the high-risk group and would have to be counseled and treated accordingly. Such calculations are a reclassification rule but not necessarily a measure of improvement. NRI or IDI for such a rule could be calculated to see how good reclassifications based on posttest risk calculations might be.

Who should be administered an additional test and economic considerations?

In the considerations presented so far, we have implicitly presented two different strategies of applying a new test.

Table 5 Diagnostic performance of the TPA in 684 originally healthy patients with 13 myocardial infarctions during follow-up (calculated from Ref. [21])

Quartiles	TPA (cm^2)	No. with AMI	10-year risk (%)	Sensitivity	Specificity	pLR	nLR
1+2	0.00–0.17	1	0.98	92	51	1.87	0.15
3	0.18–0.55	4	7.76	69	74	2.88	0.38
4	0.56–4.83	8	15.50	62	76	2.52	0.51
Best cut	1.00–4.83	8	15.50	62	87	4.77	0.44

AMI, acute myocardial infarction; nLR, negative likelihood ratio; pLR, positive likelihood ratio; TPA, total plaque area.

The first one is to add the new test to the set of standard risk factors and to refit the prediction models with the new test included. In this approach it is assumed that the new test will be performed on everyone and the data will be available for risk prediction. Some authors strongly advocate inclusion of cost considerations in the assessment of utility of new tests [24]. Their reasoning is simple: if we can assign costs to correct and incorrect diagnoses and also to testing procedures, we can come up with a decision strategy which would minimize the cost. For example, if for every patient classified as low risk who develops CHD we have cost x and for every person with high risk who does not develop an event we have cost y, the total cost can be expressed as a function of x and y weighted by the number of people in each category for a given risk threshold. If no threshold is established, averaging over all possible threshold values might be considered. Then, we can calculate the cost decrease offered by the employment of a new test by taking the difference between the costs estimated from models with and without the new test.

Proper ascertainment of costs x and y can be challenging. Moreover, costs can change over time and may not be uniform across countries. This leads to a potential problem of declaring a new test useful in one setting but not useful in another. The use of IDI and NRI provides a simple solution to this problem by assigning cost ratio of $x:y$ corresponding to the incidence ratio of the condition of interest. For example, if the incidence of CHD is 10%, $x:y$ would be 9 : 1 (missing a person who will develop CHD is nine times more costly than unnecessarily treating a person who will not develop CHD). This is arbitrary and disregards the actual (monetary) costs of diagnosis and clinical events, but the absence of more reliable estimates might be a reasonable approximation.

The second approach to calculate PTPs allows to apply the new test to a subset of individuals only, for example those who, based on the initial pretest, fall in the intermediate risk category. Such a strategy might be cost-efficient, especially when the new test is inexpensive. Of particular importance becomes the correct identification of the subpopulation that needs to be administered the

Table 6 Posttest risk calculations and 95% CI for three different levels of pretest risk, especially suited for reclassification of individuals

Pretest risk 5%			Pretest risk 10%			Pretest risk 15%		
TPA	PTP risk (%)	95% CI	TPA	PTP risk (%)	95% CI	TPA	PTP risk (%)	95% CI
0.00–0.17	0.80	0.10–5.00	0.00–0.17	1.70	0.30–10.0	0.00–0.17	2.60	0.40–15.0
0.18–0.55	8.90	7.60–10.5	0.18–0.55	18.2	14.8–19.8	0.18–0.55	**24.8**	**21.7–28.2**
0.56–4.83	11.7	7.80–17.2	0.56–4.83	**21.9**	**15.1–30.5**	0.56–4.83	**30.8**	**22.1–41.1**
1.00–4.83	**20.1**	**13.6–28.9**	1.00–4.83	**34.8**	**25.0–46.1**	1.00–4.83	**45.9**	**34.6–57.6**

Pretest risk was based on the International Atherosclerosis Society risk with an example for three individual persons. Posttest risk was calculated as exemplified in Table 4. In addition, 95% CI were calculated according to Ref. [18]. The population for posttest risk calculations were 684 originally healthy individuals with 13 myocardial infarctions during follow-up (calculated from Ref. [21], unpublished data). Bold numbers indicate high risk for coronary events. CI, confidence interval; PTP, posttest probability; TPA, total plaque area.

new (additional) test. Although this approach seems intuitive and appealing, it has potential drawbacks. It is not unlikely that if we limit the administration of the new test to the intermediate risk group, people in the low and high-risk groups who would otherwise be reclassified to a different risk group, may be missed.

Limitations

Several issues could not be covered comprehensively in this overview paper on methods about risk categorization and risk prediction. Especially questions about the ideal test on specific target populations (e.g. study cohort versus general population) or methods that more appropriately account for the increasing effect of age in elderly populations using age as the time scale. Further, many questions about cost-efficiency and cost-effectiveness would go far beyond the scope of this overview, despite their high relevance for test use appropriateness. Finally, important questions about internal validation of a new test and its external validation in independent population samples can only be mentioned within this section.

Conclusion

Prediction of cardiovascular events in originally healthy subjects remains a difficult task. Nowadays, not only do we have a large number of tests for risk prediction, either based on laboratory findings, ischemia testing or plaque imaging, but we are also confronted with a variety of statistical methods for estimating the added clinical utility of these new or emerging tests.

Based on the number of correctly identified diseased and non-diseased subjects, the performance of a test can be quantified. Whether a test is helpful or not, can be judged by a large number of criteria including pretest and PTPs, reproducibility and applicability, SE and SP, positive and NPVs. ROC analysis, successfully used to assess overall model performance, has potential deficiencies when it comes to assessing the additional independent value of a new test in medicine. It should still be calculated, but sole reliance on its effect is insufficient. In clinical practice, estimates of pretest probabilities based upon vascular risk calculators such as PROCAM, SCORE or Framingham and the subsequent calculation of PTPs based on the posttest risk calculations might be a promising approach to influence individual patient management based on true evidence and therefore clinically meaningful manner. This approach, however, is a reclassification rule and is not a measure of model performance.

Newer methods that adequately assess model performance, specifically developed for quantifying the added utility of new tests, include the NRI and IDIs. They are promising in terms of more meaningfully addressing the issue at hand, especially in comparison with c-statistics.

References

1 Riesen W, Darioli R, Noseda G, Bertel O, Buser P. Recommendations for the prevention of atherosclerosis. *Schweiz Aerztezeitung* 2005; **86**:1355–1361.
2 Von Eckardstein A, Schulte H, Assmann G. Comparison of international consensus recommendations for the detection of healthy risk subjects at high risk for myocardial infarction in Germany. *Zeitschrift für Kardiologie* 2005; **94**:52–60.
3 Conroy M, Pyörälä K, Fitzgerald AP, Sans S, Menotti A, De Backer G, *et al*. Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project. *Eur Heart J* 2003; **24**:987–1003.
4 Richter K, Lange S. Methods for the evaluation of a diagnosis. *Internist* 1997; **38**:325–336.
5 Steurer J. Expressions with high importance for the interpretation of scientific papers dealing with the description of a diagnostic test. *Schweiz Rundschau Med Praxis* 1997; **86**:945.
6 Speich R. The process of obtaining a diagnosis in internal medicine: decision analysis or intuition? Lecture at the Zurich University, April 4th 1996. *Schweiz Med Wochenschr* 1997; **127**:1263–1279.
7 Glenck U, Pewsner D, Bucher H. Evidence-based medicine: how do I evaluate a study about the value of a diagnostic test. *Schweiz Med Forum* 2001; **9**:213–220.
8 Assmann G, Cullen P, Fruchart G, Greten H, Naruszewicz M, Olsson A, *et al*. Implications of emerging risk factors for therapeutic intervention. *Nutr Metab Cardiovasc Dis* 2005; **15**:373–381.
9 Naghavi M, Falk E, Hecht H, Jamieson M, Kaul S, Berman D, *et al*. From vulnerable plaque to vulnerable patient–part III: executive summary of the screening for heart attack prevention and education (SHAPE) task force report. *Am J Cardiol* 2006; **98** (**Suppl**):2H–15H.
10 Cook N. Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation* 2007; **115**:928–935.
11 Sackett D, Strauss S, Richardson W, Rosenberg W, Haynes R. *Evidence-based medicine. How to practice and teach EBM*. 2nd ed. New York, Edinburgh, London, Toronto, Philadelphia, St. Louis, Sidney: Churchill Livingstone; 2000.
12 Ridker P, Rifai N, Rose L, Buring J, Cook N. Comparison of C-reactive protein and low-density lipoprotein cholesterol levels in the prediction of first cardiovascular events. *N Engl J Med* 2002; **347**:1557–1565.
13 O'Leary D, Wolfson S, Kronmal R, Manolio T, Burke G, Wolfson S, *et al*. Carotid-artery intima and media thickness as a risk factor for myocardial infarction and stroke in older adults. *N Engl J Med* 1999; **240**:14–22.
14 Folsom A, Chambless L, Ballantyne C, Coresh J, Heiss G, Wu K, *et al*. An assessment of incremental coronary risk prediction using C-reactive protein and other novel risk markers. The atherosclerosis risk in communities study. *Arch Intern Med* 2006; **166**:1368–1373.
15 Miller M, Zhan M, Havas S. High attributable risk of elevated C-reactive protein level to conventional coronary heart disease risk factors. The Third National Health and Nutrition Examination Survey. *Arch Intern Med* 2005; **165**:2063–2068.
16 Del Sol A, Moons K, Hollander M, Hofman A, Koudstaal P, Grobbee D, *et al*. Is carotid intima-media thickness useful in cardiovascular disease risk assessment? *Stroke* 2001; **32**:1532–1538.
17 Pencina M, D'Agostino R, D'Agostino RB Jr, Vasan R. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med* 2008; **27**:157–172.
18 Newcombe RG. Interval estimation for the difference between independent proportions: comparison of eleven methods. *Stat Med* 1998; **17**:873–890.
19 Bayes ET. An assay toward solving a problem in the doctrin of chance. *Philos Trans R Soc Lond (Biol)* 1763; **53**:370–418.
20 Fagan TJ. Nomogram for Bayes theorem. *N Engl J Med* 1975; **293**:257.
21 Martikainen J, Ottelin A, Kiviniemi V, Gylling H. Plant stanol esters are potentially cost-effective in the prevention of coronary heart disease in men: Bayesian modelling approach. *Eur J Cardiovas Prevent Rehabil* 2007; **14**:265–272.
22 Khairy P. Programmed ventricular stimulation for risk stratification in patients with tetralogy of Fallot: a Bayesian perspective. *www.nature.com/clinicalpractice*. doi:10.1038/ncpcardio0882.
23 Spence D, Eliasziw M, Di Cicco M, Hackam D, Lohmann R. Carotid plaque area: a tool for targeting and evaluating vascular preventive therapy. *Stroke* 2002; **33**:2916–2922.
24 Greenland S. The need for reorientation toward cost-effective prediction: comments on evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond by Pencina M.J. et al. *Stat Med* 2008; **27**:199–206.