

Eur. J. Clin. Chem. Clin. Biochem.
Vol. 32, 1994, pp. 169–175
© 1994 Walter de Gruyter & Co.
Berlin · New York

Performance Assessment of Coupled Tests: The Effects of Statistical Non-Independence

By A. Chiecchio¹, R. Malvano², F. Giglioli³ and A. Bo¹

¹ Servizio di Fisica Sanitaria – Ospedale Mauriziano, Torino, Italy

² Tecnostandard, Milano, Italy

³ Ospedale degli Infermi, Biella (VC), Italy

(Received August 16/December 13, 1993)

Summary: Limited to two-test associations (series and parallel schemes), the effects of statistical non-independence were studied through a mathematical approach and an experimentally-based evaluation. Both procedures were applied to results for total hormones and free fractions in euthyroid and dysthyroid subjects. Assuming independence, the sensitivity of combined tests was found to increase in parallel coupling, and to decrease, symmetrically, in series coupling, depending critically on the degree of between-test correlation and on the value of single test sensitivity (the opposite modifications obviously occur for specificity). A more complicated situation resulted for the predictive value of test associations, where a prediction based on a mathematical model was found not to be generally valid; in this case, calculations using the correct values of conditional probabilities of coupled tests seemingly remain the safest procedure.

Introduction

The increasing availability of multiple tests related to single pathologies draws attention to the assessment of the cost-effectiveness of the test associations. In this context, the non-independence of the tests is critical, as the actual values of performance characteristics (sensitivity, specificity, predictive value) can diverge from those obtained by combinatorial calculations that ignore the possible between-test correlations (1–8).

In respect to Bayes' theorem, popularised among clinical chemists by Galen & Gambino (1), different approaches have been suggested to account for the non-independence effects. Thus, it has been proposed that the conditional probabilities of single tests should be disregarded, and that decisional tables based on the test results for the evaluation of predictive value should be constructed, with a posteriori revision of the Bayesian model (6). The predictive value has also been derived from the experimental data using simplified procedures based on semiempirical formulae (7, 8). Mathematical

models have been described as well, to predict the misclassifications resulting from test correlation (2, 5).

Limited to two-test combinations, both an experimentally-based evaluation and a simplified mathematical approach have been considered in this paper, with the aim of

- (a) cross comparing the two procedures, and
- (b) evidencing the effects of the actual extent of correlations and of diagnostic features of single tests on the departure from the reference situation of independence.

For this purpose, we used as a model the data from a large number of euthyroid and dysthyroid subjects for thyroxine (T₄), free thyroxine (FT₄), triiodothyronine (T₃), free triiodothyronine (FT₃) and thyrotropin (TSH), determined by currently employed tests for thyroid status.

Materials and Methods

The groups of untreated subjects included in the study consisted of 441 euthyroids, 149 hyperthyroids and 112 hypothyroids, all with

a well defined thyroid status, as independently documented by clinical and instrumental data. All serum specimens were assayed under controlled conditions using commercial radioimmunoassay kits manufactured by Becton Dickinson (Orangeburg, NY, USA), namely T₃, T₄, FT₃ Solid Phase Component Systems and Simul-TRAC FT₄ (⁵⁷Co)/TSH MAb (¹²⁵I).

Calculations

The statistical calculations were performed with the help of a software package developed at the Ospedale Mauriziano (9), which allows:

- (a) the data distributions to be described by best-fitting probability functions;
- (b) discrimination values to be selected;
- (c) sensitivity and specificity for any test combination to be derived both from single test data on the assumption of independence and directly taking into account the position of each pair of results, related to each individual subject, with respect to the respective cutoff;
- (d) the correlation coefficient r to be calculated for all paired tests; and
- (e) all the analytes and quantities of interest to be evaluated together with their uncertainty limits (at $p \leq 0.05$, in this study), through the "bootstrap" resampling technique (10). In all the combinations, both series (AND) and parallel (OR) schemes were considered, a positive result being indicated by the outcome (++) in the former case, and by any of the outcomes (+-), (-+) and (++) in the latter.

Sensitivities and specificities for test combinations were evaluated by different procedures, as shown below.

1. Use of experimental data, without assumption of test independence, as directly provided on an experimental basis by the above mentioned program.
2. Use of model data, without assumptions of test independence, by adapting the multitest (multisymptom) mathematical approach proposed by *Norusis & Jacquiez* (2) to the case of two-test associations, according to simplified statistical formulae (11), implying functional relationships with experimental values (sample mean and standard deviation, discrimination value, correlation coefficient) evaluated as mentioned above (see Appendix A).
3. Use of combinatorial data assuming statistical independence ($r = 0$), according to the relationships

$$SE_{AND} = SE_1 \cdot SE_2, \quad SP_{AND} = SP_1 + (1 - SP_1) \cdot SP_2,$$

$$SE_{OR} = SE_1 + (1 - SE_1) \cdot SE_2, \quad SP_{OR} = SP_1 \cdot SP_2,$$

where notations SE_1 , SE_2 and SP_1 , SP_2 indicate sensitivity and specificity of the two tests, respectively, and notations SE_{AND} , SP_{AND} and SE_{OR} , SP_{OR} refer to sensitivity and specificity of the corresponding AND and OR associations.

In the opposite situation of complete correlation ($r = 1$), no gain in information will result from the test combination. When the performance of the tests differs, being $SE_1 > SE_2$ and $SP_1 > SP_2$, then the above relationship reduce to

$$SE_{AND} = SE_2, \quad SP_{AND} = SP_1,$$

$$SE_{OR} = SE_1, \quad SP_{OR} = SP_2.$$

Once sensitivity and specificity have been obtained with any of the above procedures, the predictive value of either a positive or negative result (ppv and npv, respectively) can be readily calculated from the general relationships

$$ppv = SE/[SE + pr \cdot (1 - SP)],$$

$$npv = SP/[SP + (1 - SE)/pr],$$

where pr indicates the prevalence ratio absence of disease/presence of disease.

In all cases, the data calculated on the assumption of test independence are considered (and hereafter referred to) as "reference" data.

Results and Discussion

Test correlations

Coupling of the tests in all their combinations resulted in a wide range of r values (0.0 to 0.9). Examples of between-test correlations are shown in figure 1, relative to the paired test T₃-FT₃ for both hyperthyroid and euthyroid subjects and to thyrotropin-FT₃ tests for hyperthyroid patients. These data appear to conform with a more general situation, which is encountered in the cases studied. Low correlations were invariably observed in the euthyroid group ($r = 0.0 \div 0.3$), while in dysthyroidism higher correlations resulted for all associations ($r = 0.5 \div 0.9$), excluding the thyrotropin results of the hyperthyroid group ($r = 0.0 \div 0.1$), most thyrotropin values being under the analytical limit of detection.

These findings are not unexpected. As far as euthyroid subjects are concerned, the low correlations may be explained in terms of the individual variability allowed within a non-pathological status, which is much more likely to occur than in disease, where a dynamic dimension is involved (diseased patients can aggravate and recover with time, giving rise to contiguous subpopulations). A full independence of diagnostic tests in principle conflicts with the rationale itself of their use, as the tests must correlate to some extent through their relationship with the specific pathology concerned. On the other hand, a perfect correlation expected on a clinical grounds could be weakened, or even completely hidden, by boundary conditions (e.g. the interindividual variability of carrier proteins, in the case of FT₃ and FT₄) or by the methodological characteristics themselves (e.g. the large proportions of non-measurable thyrotropin concentrations in the hyperthyroid group).

Performance assessment of test associations

The values obtained for sensitivity and specificity according to the different procedures considered are compared in table 1, relative to two examples of combined tests (T₃-FT₃ and FT₃-FT₄, euthyroidism/hyperthyroidism discrimination). In these cases, a cutoff optimization according to *Youden* (12) resulted in single test specificity as high as 95-97% and sensitivities ranging from 67 to 96%.

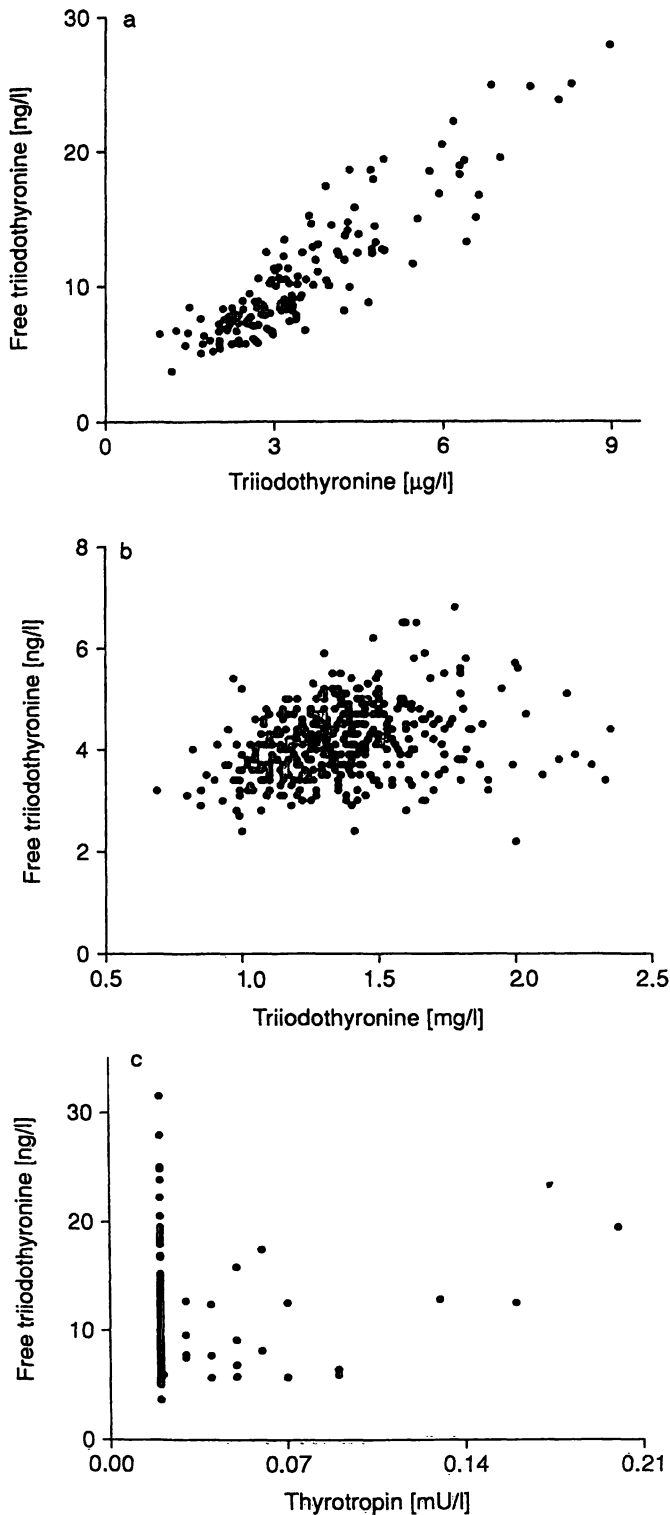


Fig. 1 Correlation of test results. (A): Hyperthyroids, $n = 149$, $r = 0.87$; (B): Euthyroids, $n = 441$, $r = 0.28$; (C): Hyperthyroids, $n = 149$, $r = 0.04$. Most thyrotropin data are under the detection limit and evaluated as 0.02 mU/l.

Some considerations emerge from these results:

- (a) in no case do significant differences exist between experimental and model data,
- (b) a general agreement is apparent for specificity values, and

Tab. 1 Evaluation of conditional probabilities for two-test associations in euthyroidism/hypothyroidism discrimination*).

	T_3 - FT_3	FT_3 - FT_4
<i>Euthyroids</i>		
r	0.28 (0.19– 0.38)	0.25 (0.17– 0.33)
Specificity %, OR:		
A	90.8	92.5
B	91.2	92.4
C	90.5 (87.8 –93.2)	92.5 (90.2 – 94.3)
AND:		
A	99.8	99.9
B	99.4	99.6
C	98.6 (97.0 –99.9)	99.8 (99.3 –100.0)
<i>Hypothyroids</i>		
r	0.81 (0.75– 0.87)	0.67 (0.54– 0.78)
Sensitivity %, OR:		
A	92.4	98.9
B	81.5	96.1
C	82.6 (75.1 –88.7)	97.3 (94.6 –100.0)
AND:		
A	51.6	73.5
B	62.4	76.2
C	61.7 (52.8 –68.0)	74.6 (66.1 – 80.1)

*) Optimization of cutoff values according to Youden (12) resulted in % sensitivities of 67.0, 77.6 and 95.5, and % specificities of 95.1, 95.0 and 97.1 for T_3 , FT_3 and FT_4 single test, respectively (mean values).

A: assumption of test independence,

B: no assumption, from mathematical model,

C: no assumption, experimental evaluation.

95% limits are indicated when provided by the computer program.

(c) in the case of sensitivity, some divergences result with respect to the reference data.

Points (b) and (c) could reflect both degree of correlation and the extent of conditional probabilities of single tests. A deeper analysis of the separate effects of these factors, as well as a confirmation of the substantial consistency of the approaches not assuming test independence, is given in figures 2 and 3 relative to sensitivity. The departure from the reference data is expressed as a function of the correlation coefficient r in the former figure, and as a function of single test sensitivity in the latter.

As a whole, these findings confirm the critical importance of considering test non-independence, in order to avoid misevaluations that would result in overestimation of the sensitivity of coupled tests in OR schemes and in underestimation in AND schemes. Obviously, the opposite situation occurs for the specificity of two-test association.

Once the correct values are available for the specificity and sensitivity of coupled tests, both ppv and npv can be properly evaluated from the general relationship previously shown (see Materials and Methods). On the other hand, the sign of the non-independence effect on ppv and npv (increase or decrease with respect to reference) could not be anticipated, since the values of the

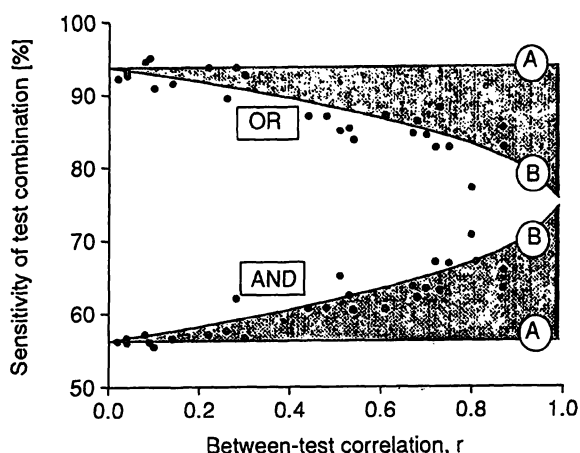


Fig. 2 Effect of test correlation on the departure of the actual values of association sensitivity from those calculated on the assumption of test independence. Correlation is expressed as absolute *r* values, ignoring the algebraic sign. To simulate the sensitivities of weakly correlated associations, the specificity data were actually used (in this case, the AND scheme was assumed as coinciding with the unanimity rule for both pathological and non-pathological samples, considering as positive results the (++) and (--) outcomes, respectively). A 75% sensitivity (or specificity) was taken for single tests. The values obtained with combinatorial calculation assuming independence (A) and those provided by the predictive mathematical model (B) are expressed as continuous lines. The mean values obtained experimentally are shown by symbols, omitting the uncertainty ranges for graphical simplicity.

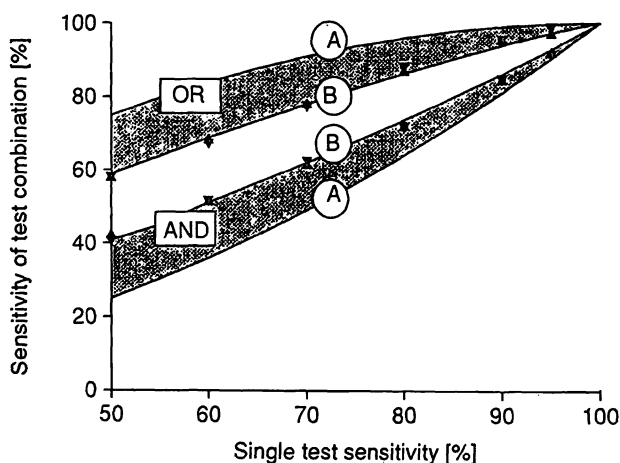


Fig. 3 Effect of the sensitivity of single tests (assumed as equivalent) on the departure of the actual values of association sensitivity from those calculated on the assumption of test independence. The data refer to the combination of triiodothyronine-free triiodothyronine tests. The values of the correlation coefficient *r* used in the mathematical model is 0.84, the mean experimental values being 0.87 and 0.81 in the case of hyperthyroid (▼) and hypothyroid (▲) groups, respectively. For other details, see the caption of figure 2.

variables involved in these relationships (SE and 1-SP, SP and 1-SE) are modified in the same direction to extents depending on any given situation.

A simplification should in principle derive from the low test correlations observed in the absence of disease, which in turn entail the absence of significant non-inde-

pendence effects on specificity. Constant SP and 1-SP values should allow a prediction of the sign of departure from the reference data of the actual predictive values, i. e. with respect to the reference situation: higher values in the AND scheme and lower values in the OR scheme for both ppv and npv.

The effects of non-independence on the predictive value of test associations were experimentally evaluated for the T₃, FT₃, T₄ and FT₄ tests coupled in all their combinations. These associations could possibly reflect a more general situation, as any methodological effect on the actual test correlations is absent, in the non-pathological state the correlations are low but still significant ($r = 0.2 \div 0.3$), while they are relatively high in disease ($r = 0.6 \div 0.9$). The examples given in tables 2 and 3 for ppv

Tab. 2 Evaluation of the positive predictive value of the T₃-FT₃ association in euthyroidism/hypothyroidism discrimination*).

Prevalence ratio euthyroidism/hypothyroidism	Evaluation approach	Predictive value [%]	
		AND	OR
81	A	73.5	11.6
	B	51.0	10.0
	C	35.1 (22.6-62.4)	10.0 (7.7-13.7)
27	A	90.3	27.5
	B	79.2	26.0
	C	64.9 (45.4-84.2)	25.6 (18.8-31.5)
9	A	96.1	49.5
	B	90.4	48.2
	C	82.1 (69.2-92.2)	47.0 (39.4-53.3)
3	A	99.0	78.5
	B	97.1	78.1
	C	94.6 (89.5-97.2)	78.1 (69.8-81.8)

*) See footnote to table 1.

Tab. 3 Evaluation of the negative predictive value of the T₃-FT₃ association in euthyroidism/hypothyroidism discrimination*).

Prevalence ratio euthyroidism/hypothyroidism	Evaluation approach	Predictive value [%]	
		AND	OR
27	A	98.2	99.2
	B	98.6	99.2
	C	98.6 (98.2-98.9)	99.2 (98.9-99.5)
9	A	94.9	99.1
	B	96.0	97.8
	C	95.9 (94.9-96.8)	97.9 (97.1-98.7)
3	A	86.1	97.3
	B	88.8	93.8
	C	88.7 (85.9-90.7)	94.0 (91.5-96.2)
1	A	67.4	92.2
	B	72.6	83.4
	C	72.2 (67.4-77.0)	83.9 (78.2-89.2)

*) See footnote to table 1.

and npv, respectively, relative to the combined T_3 and FT_3 tests (euthyroidism/hypothyroidism discrimination) are perfectly representative of the overall set of experimental data.

A scrutiny of the results has led to the following observations:

a) Surprisingly enough, contrary to expectations, the ppv for AND coupling constantly resulted in values lower than the reference data (the role of prevalence being expectedly critical in increasing the differences).

b) Non-independence proved substantially ineffective in the case of ppn in OR schemes; some fluctuations around the reference values (resulting in a slight decrease of ppv in the example of table 2) are seemingly due to the uncertainty associated with the variables concerned.

c) In correspondence with low disease prevalences, non-independence was also found to be ineffective for npv, in both AND and OR schemes. As expected, however, (exceptionally) high disease prevalences were accompanied by increased values in AND associations and decreased values in OR associations.

d) The data provided by the mathematical model were found to coincide with those experimentally obtained in the case of npv and, obviously, of ppv in OR coupling, while systematic discrepancies resulted for ppv in AND schemes (the model values being intermediate between mean experimental values and reference data).

The last point may be explained by the inadequacy of the average r values evaluated in the euthyroid group for correctly describing the correlation in the euthyroid tail regions (1-SP), while they always proved adequate in the case of dysthyroid groups. This invalidates the mathematical approach where ppv for AND coupling is concerned, as properly fitting r values cannot be independently established.

The formulas extending the mathematical model to the case of predictive value (see Appendix) can elucidate the behaviour shown by ppv. Conformity with expectation in AND schemes, implying the inversion of the differences observed, should require either unrealistic conditions (such as single test specificities or sensitivities approaching zero), or conditions which are possibly infrequent, and in any case apparently not met within the experimental situation considered (such as very low correlations in the tail regions of non-pathological values) (see Appendix B). As for OR coupling, the tendency of ppv to remain unchanged is explained by the fact that the only effective factor in the relationship concerned (see Appendix B) relies on the essentially stable test specificity.

Conclusions

The widespread availability of computing facilities in clinical chemistry laboratories greatly helps in establishing diagnostic strategies and cost/benefit analyses, without limiting or misleading assumptions, such as the statistical independence of tests. From this point of view, therefore, the use of either experimental approaches (as dealt with in this paper) or mathematical models for the definition of sensitivity and specificity of combined tests are per se equally suitable. Arguments in favour of the former procedure could be, in principle, the absence of any working hypothesis, as well as the ease of defining variability ranges more adequate to actual situations.

On practical grounds, the readily measurable values of the between-test correlation coefficient and/or conditional probabilities of the single tests could help in predicting whether or not the independence hypothesis may be accepted in test coupling. A rule of thumb could be the rejection of the hypothesis for, e.g. r values higher than 0.3 and/or single test sensitivities and specificities lower than 95% (provided that a sufficient number of results, e.g. 100 or more, are available for each test and clinical situation). Though possibly not a valid generalization, the low correlation observed for non-pathological cases – and hence the adequacy of the assumption of test independence – could be suggestive of less critical effects in assessing the specificity of test combinations.

As for the predictive values of the results of coupled tests, the situation appears to be more complicated than suggested by the literature (3, 6–8), and direct evaluations through mathematical models are not generally valid. A calculation using the correct values of conditional probabilities of test associations (obtained by any means) seemingly remains the safest and easiest way to follow.

Appendix

A. The random variables (the tests) X and Y are assumed to distribute as a bivariate normal distribution, with means m_x and m_y and variances s_x^2 and s_y^2 , respectively, and an expected correlation coefficient ρ . If the test value $x \in X$ is less than (or equal to) h , and $y \in Y$ is less than (or equal to) k , then the joint probability Pr is given by

$$(1) \quad Pr\{X \leq h_0, Y \leq k_0\} \\ = \frac{1}{s_x s_y} \int_{-\infty}^{(h_0 - m_x)/s_x} \int_{-\infty}^{(k_0 - m_y)/s_y} g(s, t, \rho_{\Omega}) ds dt \\ = L\left(-\left(\frac{h_0 - m_x}{s_x}\right), -\left(\frac{k_0 - m_y}{s_y}\right), \rho_{\Omega}\right) = L(h_{\Omega}, k_{\Omega}, \rho_{\Omega})$$

where

$$g(x, y, \rho) = \frac{1}{2\pi(1-\rho^2)} e^{-\frac{1}{2}\left(\frac{x^2 - 2\rho xy + y^2}{1-\rho^2}\right)},$$

and

$$(2) L(a, b, \rho) = Q(a)Q(b) + \sum_{n=0}^{\infty} \frac{Z^n(a)Z^n(b)}{(n+1)!} \rho^{n+1},$$

being

$$Q(x) = \frac{1}{2\pi} \int_x^{+\infty} e^{-\frac{s^2}{2}} ds, Z^n(x) = \frac{1}{2\pi} \frac{d^n}{dx^n} Z(x), Z(x) = \frac{1}{2\pi} e^{-\frac{x^2}{2}},$$

and h_0, k_0 , of the situation in study (Positive or Negative).

In the present study, the formula (2) was calculated using an approximation, such that the difference between the exact value and the actual value does not exceed 0.1%.

When the random variables X and Y are not normally distributed, the applicability of the formulae shown above and in section B requires a preliminary standardization of data, such as that reported in l. c. (9).

B. When relationships (1) and (2) are applied to the case of the positive predictive value for AND and OR coupling, respectively, then the general formulae

$$(3_{AND}) \quad ppv_{AND} = \frac{1}{1 + pr \frac{1 - SP_{AND}}{SE_{AND}}}$$

$$(3_{OR}) \quad ppv_{OR} = \frac{1}{1 + pr \frac{1 - SP_{OR}}{SE_{OR}}}$$

can be written as

$$(4_{AND}) \quad ppv_{AND} = \frac{1}{1 + pr \frac{L(h_{SP_{AND}}, k_{SP_{AND}}, \rho_{SP})}{L(-h_{SE_{AND}}, -k_{SE_{AND}}, \rho_{SE})}}$$

$$(4_{OR}) \quad ppv_{OR} = \frac{1}{1 + pr \frac{1 - L(-h_{SP_{OR}}, -k_{SP_{OR}}, \rho_{SP})}{1 - L(h_{SE_{OR}}, k_{SE_{OR}}, \rho_{SE})}}$$

where SP and SE refer to the non-pathological and the pathological status, respectively, and pr indicates the prevalence ratio *absence of disease/presence of disease*. Rewriting (4_{AND}) and (4_{OR}) by using a first-order approximation of (2), the following relationships are obtained

$$(5_{AND}) \quad ppv_{AND} = \frac{1}{1 + pr \frac{Q(h_{SP_{AND}})Q(k_{SP_{AND}}) + \rho_{SP}Z(h_{SP_{AND}})Z(k_{SP_{AND}})}{Q(-h_{SE_{AND}})Q(-k_{SE_{AND}}) + \rho_{SE}Z(h_{SE_{AND}})Z(k_{SE_{AND}})}} = \frac{1}{1 + pr \frac{(1 - SP_h)(1 - SP_k) + \rho_{SP}Z(h_{SP_{AND}})Z(k_{SP_{AND}})}{SE_h SE_k + \rho_{SE}Z(h_{SE_{AND}})Z(k_{SE_{AND}})}}$$

$$(5_{OR}) \quad ppv_{OR} = \frac{1}{1 + pr \frac{1 - Q(-h_{SP_{OR}})Q(-k_{SP_{OR}}) - \rho_{SP}Z(h_{SP_{OR}})Z(k_{SP_{OR}})}{1 - Q(h_{SE_{OR}})Q(k_{SE_{OR}}) - \rho_{SE}Z(h_{SE_{OR}})Z(k_{SE_{OR}})}} = \frac{1}{1 + pr \frac{1 - SP_h SP_k - \rho_{SP}Z(h_{SP_{OR}})Z(k_{SP_{OR}})}{1 - (1 - SE_h)(1 - SE_k) - \rho_{SE}Z(h_{SE_{OR}})Z(k_{SE_{OR}})}}$$

For independent tests, equations (3_{AND}) and (3_{OR}) become

$$(6_{AND}) \quad \overline{ppv}_{AND} = \frac{1}{1 + pr \frac{(1 - SP_h)(1 - SP_k)}{SE_h SE_k}}$$

$$(6_{OR}) \quad \overline{ppv}_{OR} = \frac{1}{1 + pr \frac{1 - SP_h SP_k}{1 - (1 - SE_h)(1 - SE_k)}}$$

As it is expected that (5_{AND}) ≥ (6_{AND}), then

$$(7_{AND}) \quad \rho_{SP} \leq \rho_{SE} \frac{(1 - SP_h)(1 - SP_k)}{SE_h SE_k} \frac{Z(h_{SE_{AND}})Z(k_{SE_{AND}})}{Z(h_{SP_{AND}})Z(k_{SP_{AND}})}$$

This inequality is satisfied when SP_h, SP_k, SE_h and SE_k approximate 0.0%

(note that $\lim_{\alpha \rightarrow 0} \left(\frac{1 - \alpha}{Z(h_\alpha)}\right) = +\infty$ and $\left(\frac{Z(h_\alpha)}{\alpha}\right) = \max$ for $\alpha \cong 0$).

On the contrary, it is expected that (6_{OR}) ≥ (5_{OR}). Then

$$(7_{OR}) \quad \rho_{SP} \leq \rho_{SE} \frac{1 - SP_h SP_k}{1 - (1 - SE_h)(1 - SE_k)} \frac{Z(h_{SE_{AND}})Z(k_{SE_{AND}})}{Z(h_{SP_{AND}})Z(k_{SP_{AND}})}$$

and if $1 - (1 - SE_h)(1 - SE_k) \cong 1$, as occurs in practice, then

$$(7_{OR})^* \quad \rho_{SP} \leq \rho_{SE} (1 - SP_h SP_k) \frac{Z(h_{SE_{AND}})Z(k_{SE_{AND}})}{Z(h_{SP_{AND}})Z(k_{SP_{AND}})}$$

In this situation, under the same conditions, (7_{OR})^{*} is less demanding than (7_{AND}) (e. g. of a factor ca 20, for values around 90%, of both SP_h and SP_k).

References

1. Galen, R. S. & Gambino, S. R. (1975) *Beyond Normality: The Predictive Value and Efficiency of Medical Diagnoses*. Wiley, New York.
2. Norusis, M. J. & Jaquiez, J. A. (1975) Diagnosis I: Assumption of nonindependence in mathematical models for diagnosis. *Comput. Biomed. Res.* 8, 156–172.
3. Cebul, R. D., Hershey, J. C. & Williams, S. V. (1982) Using multiple tests: Series and parallel approaches. *Clinics in Laboratory Medicine* 2, 871–890.
4. Chinchilli, V. M. (1983) Estimation of sensitivity and specificity in a multistage screen for medical diagnosis. *Biometrics* 39, 333–340.
5. Russek, E., Kronmal, R. A. & Fisher, L. D. (1983) The effects of assuming independence in applying Bayes theorem to risk estimation and classification in diagnosis. *Comput. Biomed. Res.* 16, 537–552.
6. Jones, R. H. & McClatchey, M. W. (1988) Beyond sensitivity, specificity, and statistical independence. *Stat. Med.* 7, 1289–1295.
7. Marshall, R. J. (1989) The predictive value of simple rules for combining two diagnostic tests. *Biometrics* 45, 1213–1222.
8. Borowiak, D. & Reed, J. F. (1991) Utility of combining two diagnostic tests. *Comput. Meth. Programs Biomed.* 35, 171–175.
9. Chiecchio, A., Bo, A., Manzone, P. & Giglioli, F. (1993) DECIDE: A software for computer-assisted evaluation of diagnostic test performance. *Comput. Meth. Programs Biomed.* 40, 55–65.
10. Efron, B. (1982) *The jackknife, the bootstrap and other resampling plans*. SIAM, Philadelphia.
11. Zelen, M. & Severo, N. (1970) Probability functions. In: *Handbook of Mathematical Functions* (Abramowitz, M. & Stegun, I. A., eds.) pp. 925–995, Dower Publications, New York.
12. Youden, W. J. (1950) Index for rating diagnostic tests. *Cancer* 3, 32–35.

Dr A. Chiecchio
Servizio di Fisica Sanitaria
Ospedale Mauriziano
Largo Turati, 62
I-10128 Torino
Italy

