The radiological diagnosis in the light of Information Theory: a perspective.

Poster No.:	C-0505
Congress:	ECR 2015
Type:	Educational Exhibit
Authors:	R. Girometti[1], F. Fabris[2]; [1]Udine/IT, [2]Trieste/IT
Keywords:	Education and training, Statistics, Education, Experimental, Professional issues, Management, Computer applications
DOI:	10.1594/ecr2015/C-0505

Learning objectives

1. Radiologists can be easily familiar with: a) main concepts of Information Theory (IT); b) measures of information defined by IT.

2. Radiological diagnosis can be modeled in terms of IT, providing informational measures of the accuracy of a diagnostic test.

3. Informational measures have the potential to complement and refine standard analysis of accuracy.

Background

Standard analysis of accuracy

Standard analysis (SA) of the accuracy of a diagnostic test is based on the comparison between the test result and a standard of reference (e.g., histological examination), that is on the probabilities that a test really represents the presence or absence of a certain medical condition [1] (Fig. 1 on page 3).

However, SA shows the limitations illustrated in Fig. 2 on page 4 [1-2], including the lack of a rigorous quantification of the informational content underlying the diagnostic process.

Basic questions

1. Can we assess the performance of a diagnostic test in informational terms? In other words, can we measure whether a test conveys a reasonable amount of information for performing a diagnosis and/or managing a patient in a certain clinical scenario?

2. Can informational analysis provide summary measures of the accuracy of a diagnostic test?

The answer is: yes, by using measures of information from the Information Theory (IT).

Information Theory

Foundations of the IT have been built in 1948 by the American mathematician and engineer C.E. Shannon (1916-2001). IT formalizes the mathematical rules underlying

telecommunications and defines the general properties that communication systems should have to transmit information reliably and affordably [3]. Not surprisingly, IT is a pillar in telecommunication technologies [4].

In order to define rigorous and objective measures of information, Shannon formalized a general model of communications systems, namely the "binary transmission channel model", where the "channel" is whatever medium used to transmit a signal from a transmitter to a receiver. Basic properties of the transmission channel are shown in Fig. 3 on page 5 .

The measure of the information flowing through the channel is named Mutual Information (*MI*). MI can be defined as the average quantity of information *I(X,Y)* that two variables *X* and *Y* exchange each other (e.g., the input and output in the binary transmission channel model) [5-6]. Properties of *MI* are summarized in Fig. 4 on page 6 .

Images for this section:

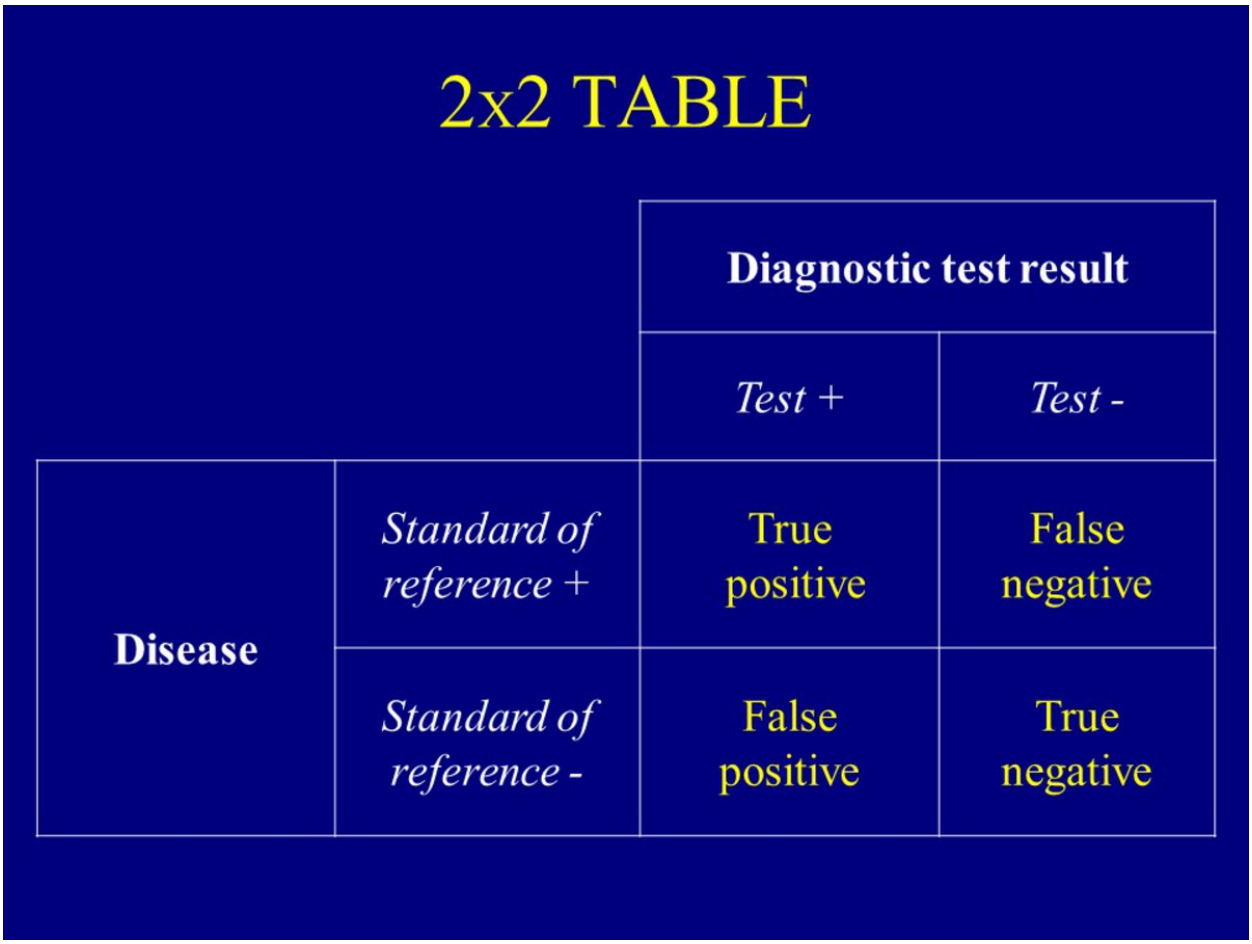


Fig. 1: Conventional statistical approach to diagnostic accuracy is based on the comparison between the results of a test and a standard of reference, as exemplified by the classical 2x2 table.

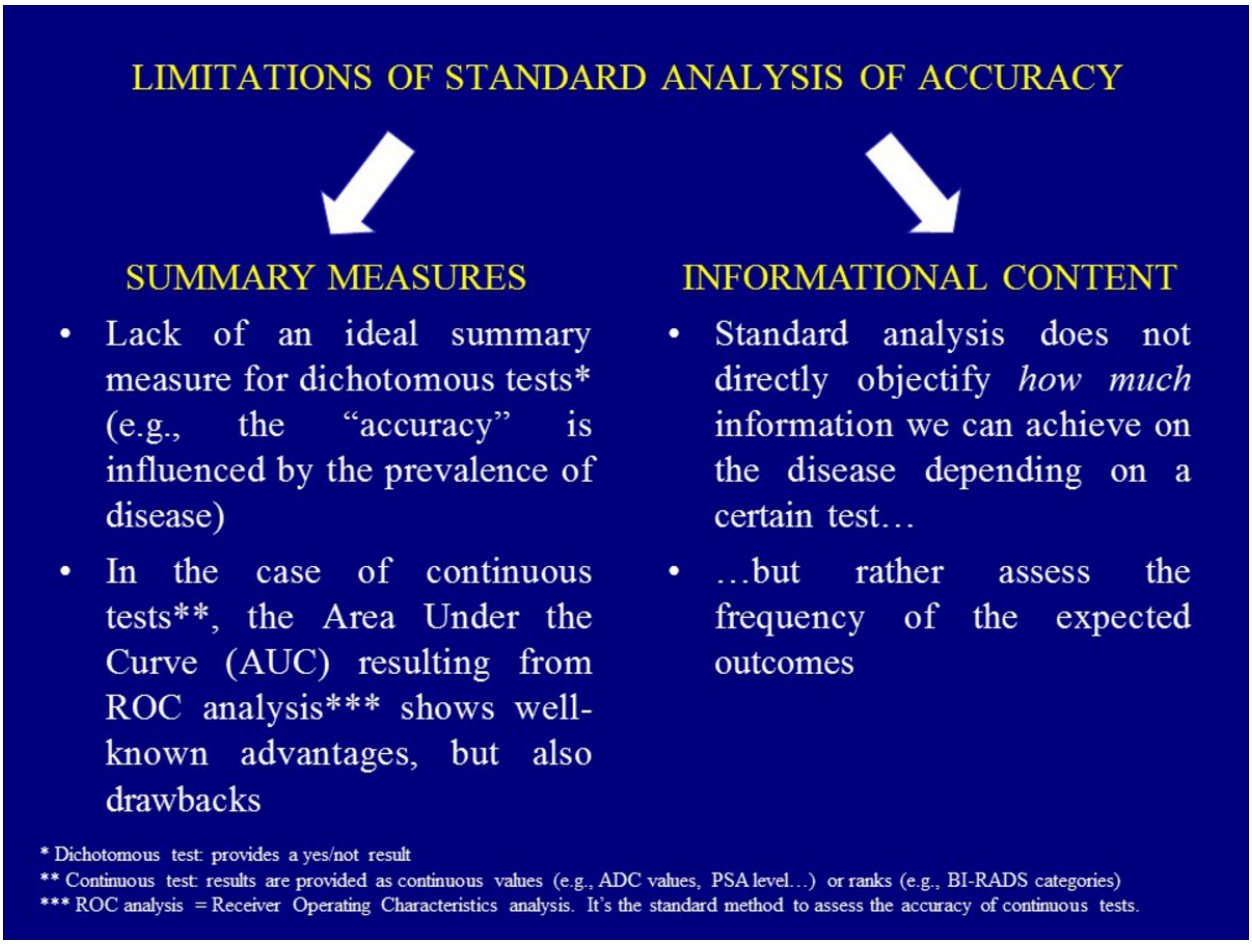


Fig. 2: Limitations of standard analysis of diagnostic accuracy.

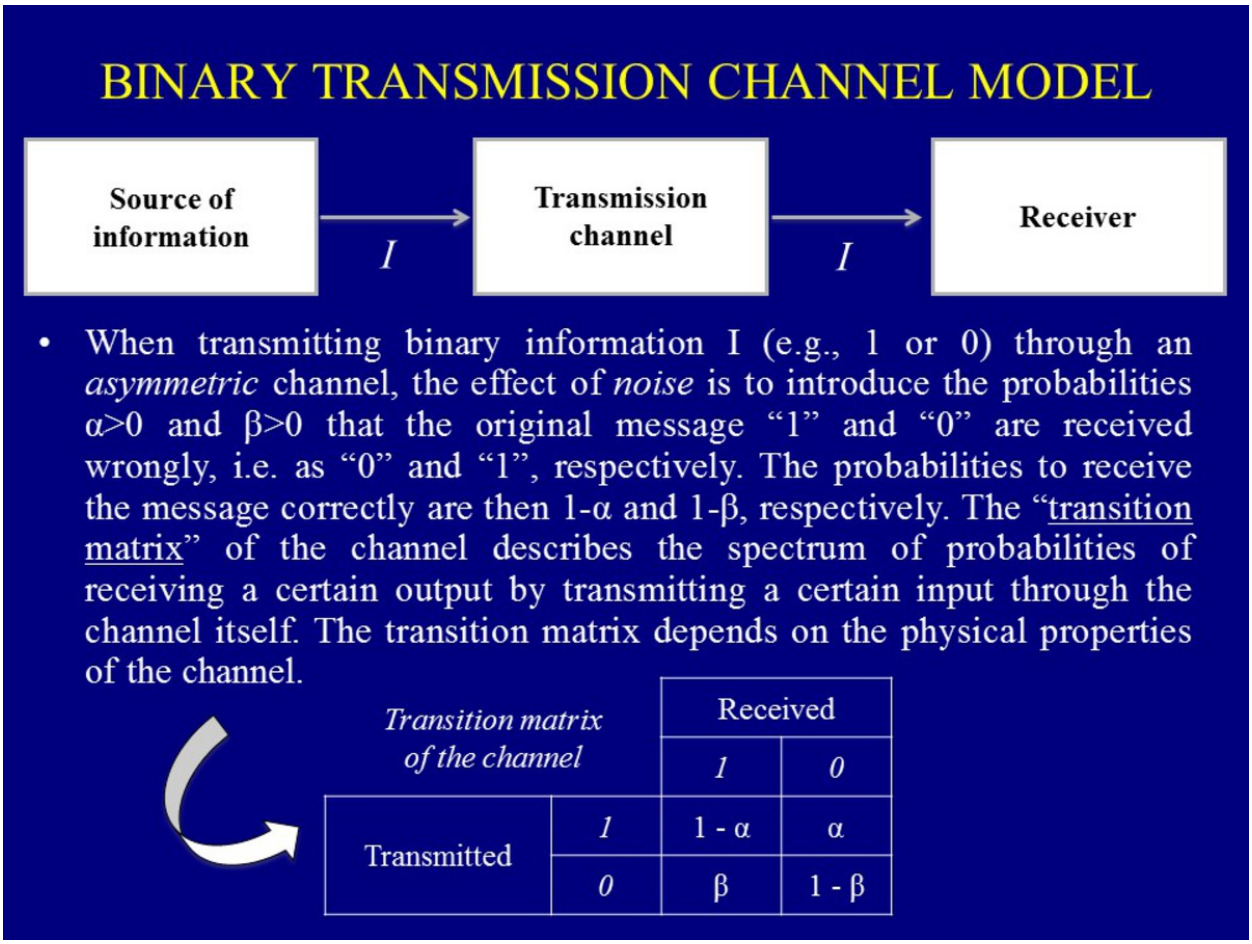
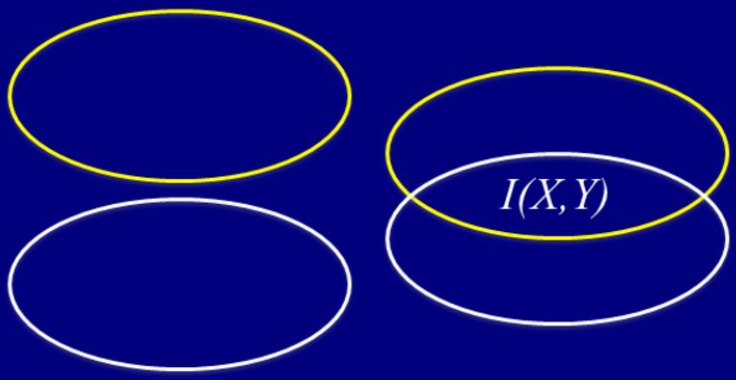


Fig. 3: Basic properties of a «binary transmission channel system».

MUTUAL INFORMATION

Probability distribution of the events $x_1, x_2, \ldots, x_i$ (input in the channel)

Probability distribution of events $y_1, y_2, \ldots, y_i$ (output from the channel)



Are variables independent?	Yes, variables are not correlated	No, variables show a certain degree of correlation
Do variables share information?	No (knowing X gives no information about Y, and vice versa)	Yes (knowing X gives some Information on Y, and vice versa)
$MI = I(X,Y)$	0	> 0

Fig. 4: Properties of Mutual Information.

Findings and procedure details

The diagnostic channel

How can we express radiological diagnosis with IT? By modeling the radiological examination as an asymmetric "diagnostic channel" [3-6]. In this model, information flows from the source (the disease) to the receiver (the radiologist) through an imaging tool (the diagnostic channel = CT, MRI, mammography and so on…). It is easy to demonstrate that the probabilities of correct/incorrect transmission of the signal "disease" (i.e., the transition matrix of the diagnostic channel) are equivalent to the probabilities calculated when cross tabulating test results vs. the standard of reference results (Fig. 5 on page 10).

Definition of "accuracy" in informational terms (in three steps)

1. In the above model, a test is as more accurate as more information on the disease is prompted to the radiologist, that is**:** the larger the MI, the larger the diagnostic information *I(D,R)* exchanged between the disease (D) and the radiologist (R). In mathematical terms, this translates into (see Appendix I):

$$I(D,R) = PREV \cdot SE \cdot \log SE/PR + PREV \cdot FNR \cdot \log FNR/NR + (1\text{-}PREV) \cdot FPR \cdot \log FPR/PR + (1\text{-}PREV) \cdot SP \cdot \log SP/NR$$

(1),

where PREV = prevalence of the disease in the study population, i.e. the pre-test probability of disease; SE = sensitivity; SP = specificity; PR and NR = probability of a positive or negative test result, respectively. The equation tells us that MI depends on the pre-test probability of disease and the transition matrix of the channel.

2. To build a proper measure of accuracy, we must define now the "Capacity" of the (binary asymmetric) channel (C_{BAC}), that is the maximum amount of information *I(D,R)* flowing through the diagnostic channel, given all possible PREV values (see Appendix II):

$$C_{BAC} = \max MI \text{ over all PREV values} \quad (2).$$

C_{BAC} corresponds to the maximum accuracy of a test.

3. Since we don't know in advance whether the PREV in the study population is the one corresponding to C_{BAC}, informational accuracy should be intended globally as the MI-

curve, i.e. the area under the curve (AUC) obtained by plotting MI versus all possible pre-test probabilities of disease (Fig. 6 on page 11) (see Appendix III).

Measures of informational accuracy

The MI-curve is the base to derive summary measures of informational accuracy, namely the Information ratio (*IR*) for dichotomous tests and the Global Information Ratio (*GIR*) for continuous tests, respectively.

1. The IR *IR* is the ratio between the *MI*-curve of a test and the *MI*-curve of the standard of reference (SR) (that plots *MI* variations over the *PREV* when, by definition, SE = 1.0, SP = 1.0, TNR = 0 and FPR = 0) (Fig. 7 on page 12). *IR* is a dimensionless value expressing, on a range between 0 and 1, how much of the maximum possible information about the disease the test is able to convey (Fig. 8 on page 13).

2. In the case of tests expressing the results on an interval or rank scale (e.g., ADC values, BI-RADS…) we propose, similarly to ROC analysis, to calculate the *IR* value for each of the 2x2 tables obtained by varying the threshold # = 1- SP. This leads to: A) calculate which threshold corresponds to the maximum IR; B) build an Information Ratio Curve (IRC) and calculate the GIR as a summary measure of accuracy (Fig. 9 on page 14). The GIR expresses, on a range between 0 and 1, how much of the maximum possible information about the disease the test is able to convey. (See Appendix IV)

Application to a real scenario

REFERENCE STUDY

In order to test the model, we applied it to real data from a celebrated work by Pisano et al. [7], comparing Digital Mammography (DM) vs. Film Mammography (FM) in diagnosing breast cancer in a screening population of 42,760 women recruited over 33 referral centers. The Authors assessed the accuracy of DM and FM with ROC analysis on the entire study group and several subgroups, using a seven-points scale to classify mammography findings: 1 = definitely not malignant; 2 = almost definitely not malignant; 3 = probably not malignant; 4 = possibly malignant; 5 = probably malignant; 6 = almost definitely malignant; 7 = definitely malignant). The standard of reference was breast biopsy performed within 455 days after the study entry, and/or a follow-up mammogram obtained at last 12 months after the study entry. In the analysis, scores 4 to 7 were defined as positive results, while score 1 to 3 were defined as negative results.

We tested the informational model on data showed in Table 3 of the original paper after recalculating unfitted, empirical operating points [8] of the ROC curve using the 7-point scale on the entire study cohort after 455 days of follow-up.

RESULTS AND COMMENTS

1. As shown in Table 1 on page 15, the cut-off of maximum *IR* for FM corresponded to the best cut-off of SE/1-SP of ROC analysis. On the other hand, maximum *IR* for DM corresponded to a lower cut-off compared to the one associated with best SE/1-SP at ROC analysis. Our explanation for this finding is that additional information provided by DM translates into better visibility of subtler anatomic details mimicking malignancy (e.g., spicules, microcalcifications…). Thus, complementing ROC analysis with informational analysis might be of help in understanding "how" information associated to novel technologies impacts on readers in certain clinical scenarios.

2. We found larger percentage increment in accuracy (DM vs. FM) as assessed by *GIR* (+ 14.5%) than ROC analysis (+ 2.45%), suggesting that informational analysis better emphasized the difference in performance between imaging modalities (Fig. 10 on page 16).

Final remarks

1. The MI-curve tends to zero when the pre-test probability of disease approaches 0 or 1, since a test applied to a population where all the members are ill (or healthy) carries no information. The maximum amount of information is associated to a pre-test probability close to ½.

2. IR results from a MI-curve representing the MI values over all possible pre-test probabilities of disease. Consequently, IR is independent from the pre-test probability in a given clinical experiment, expressing the maximum information for all clinical scenarios at one time.

Mathematical Appendix

Here are some slides with mathematical details underlying the above concepts (Fig. 11 on page 17, Fig. 12 on page 18, Fig. 13 on page 19, Fig. 14 on page 20).

Images for this section:

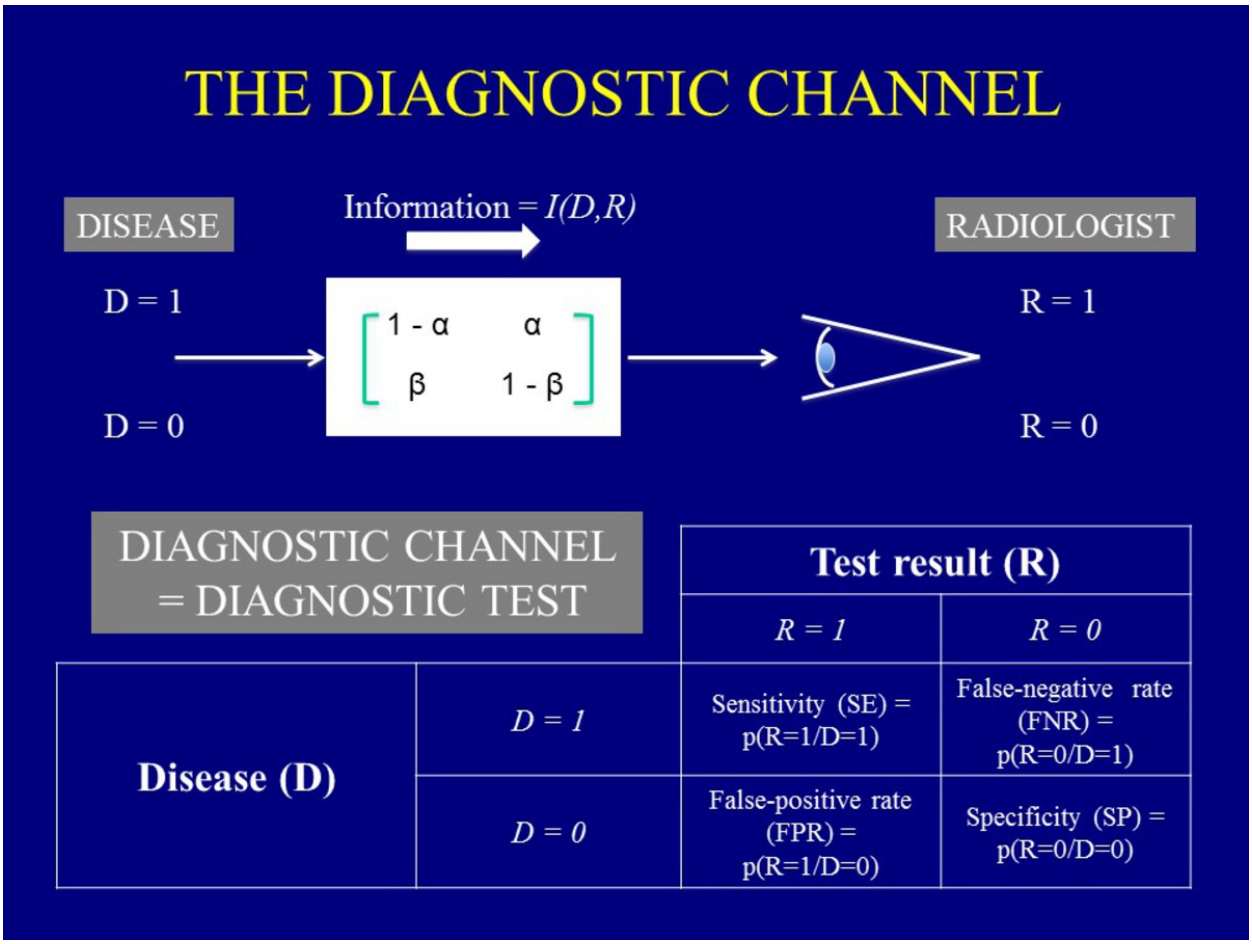


Fig. 5: Properties of a diagnostic tool as a «diagnostic channel». We chose the asymmetric binary channel model (in which # # #), because false-positives and false-negatives (errors in trasmission) do no strictly have the same probability to occur (see Fig. 3). In IT, the most widely used model is that of the «symmetric» channel, in which # = # [6]. PLEASE NOTE: SE, SP, FNR and FPR are expressed under the form of conditional probabilities (e.g., sensitivity is the probability of "R positive" given "D present").

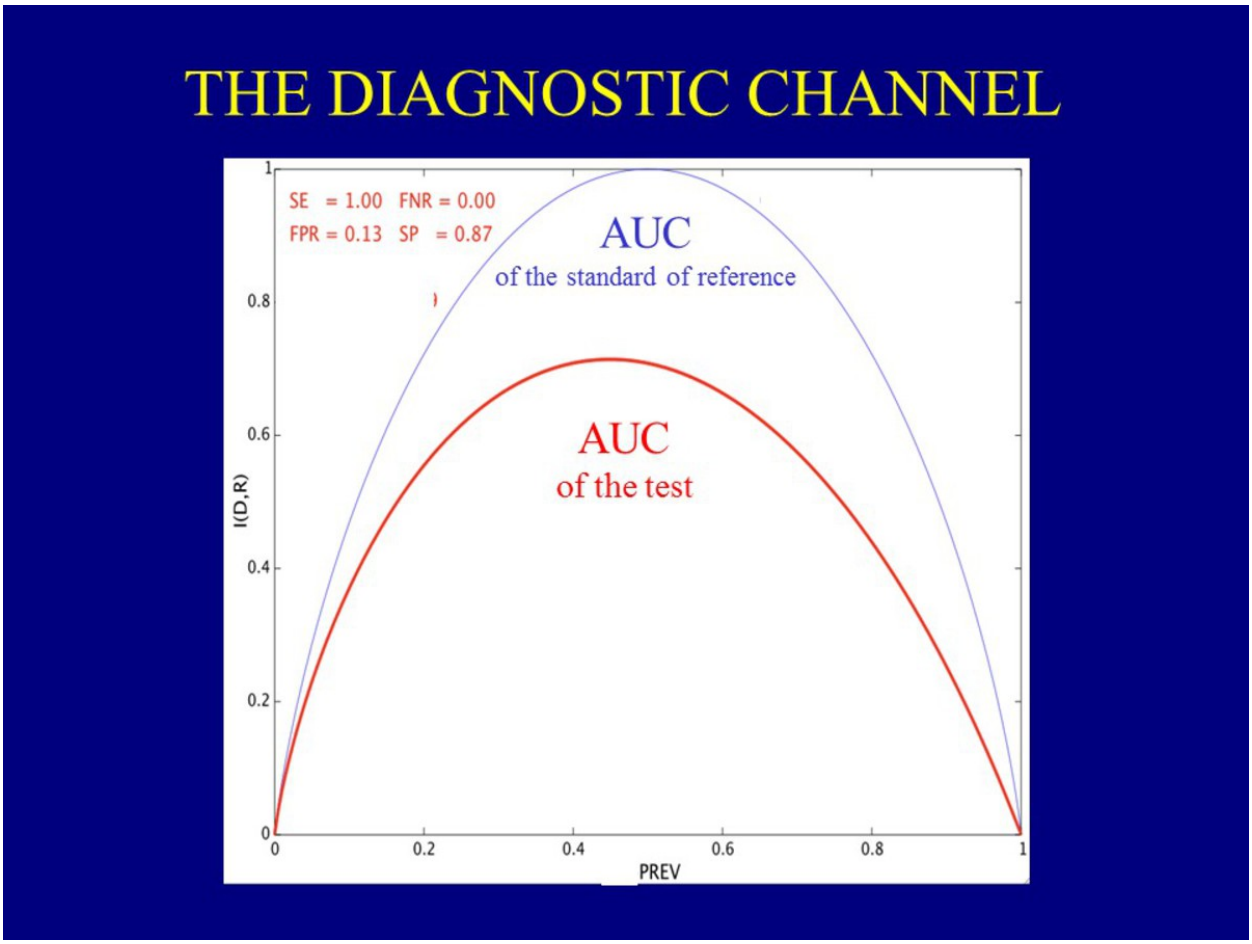


Fig. 6: MI-Curves obtained by plotting the MI values I(D,R) of a test versus the pre-test probability of disease. The blue curve refers to the MI-curve of the standard of reference (i.e., a test that is always correct), whereas the red curve refers to the MI-curve of a test showing reported sensitivity and specificity. AUCs represent global accuracy of a test.

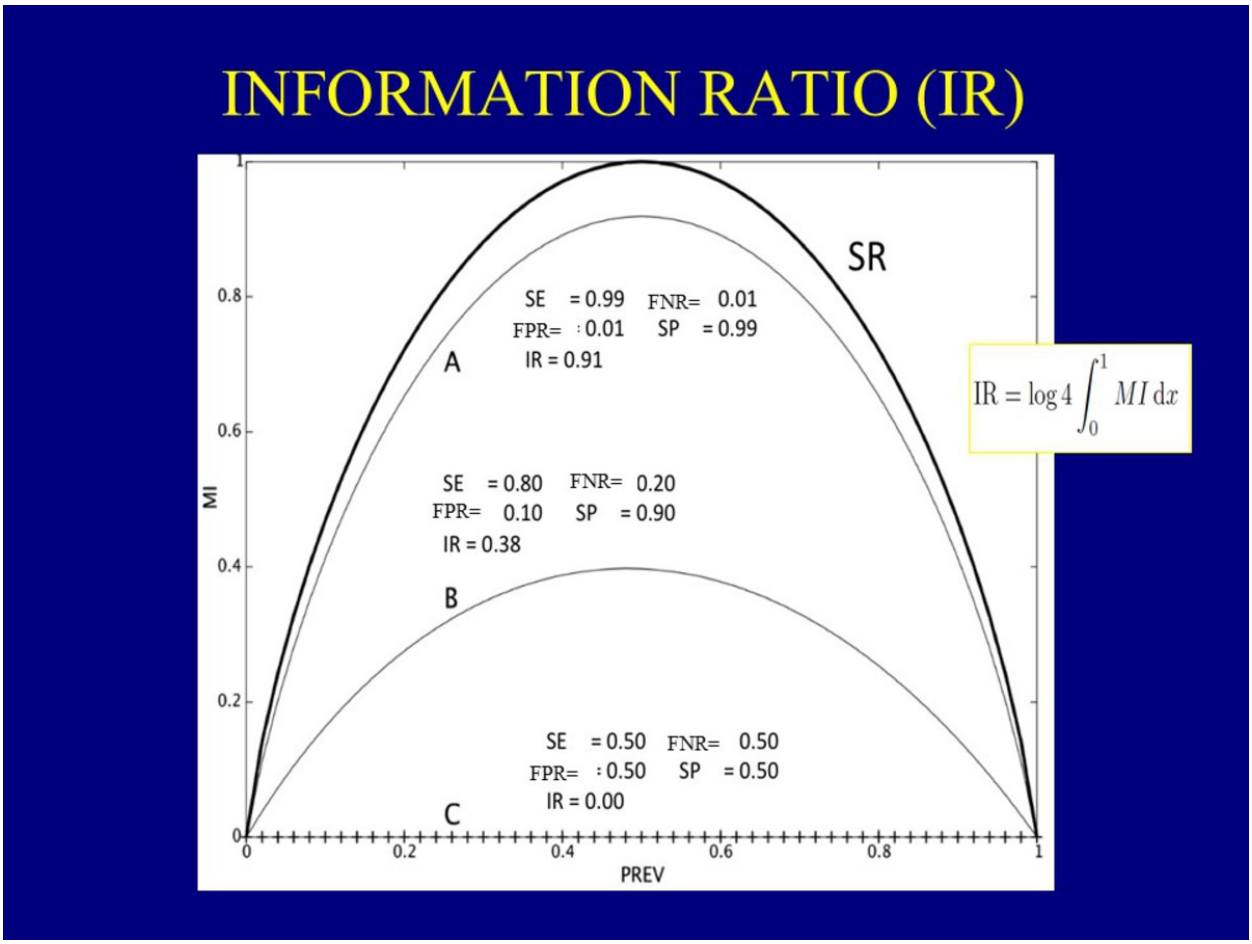


Fig. 7: Information associated to three exemplificative tests A, B and C as compared to the standard of reference (SR). SE = sensitivity; SP = specificity; FNR = false negatives rate; FPR = false positives rate; IR = Information ratio.

INFORMATION RATIO (IR)

- E.g., for a diagnostic test with $\alpha = \beta = 0.5$, the accuracy is 50%, whereas *IR* is 0
- In other words, there is no information flowing through the diagnostic channel, since by definition, saying $\alpha = \beta = 0.5$ is equivalent to choice the diagnosis randomly

$\alpha = \beta$	*IR* %	Accuracy %
0	100.00	100.00
0.005	94.68	99.50
0.01	90.72	99.00
0.02	84.11	98.00
0.05	68.92	95.00
0.1	50.56	90.00
0.2	26.06	80.00
0.3	11.03	70.00
0.4	2.69	60.00
0.45	0.67	55.00
0.5	0.00	50.00

** Accuracy = (TP+TN)/(TP+TN+FP+FN)*

Fig. 8: Properties of IR.

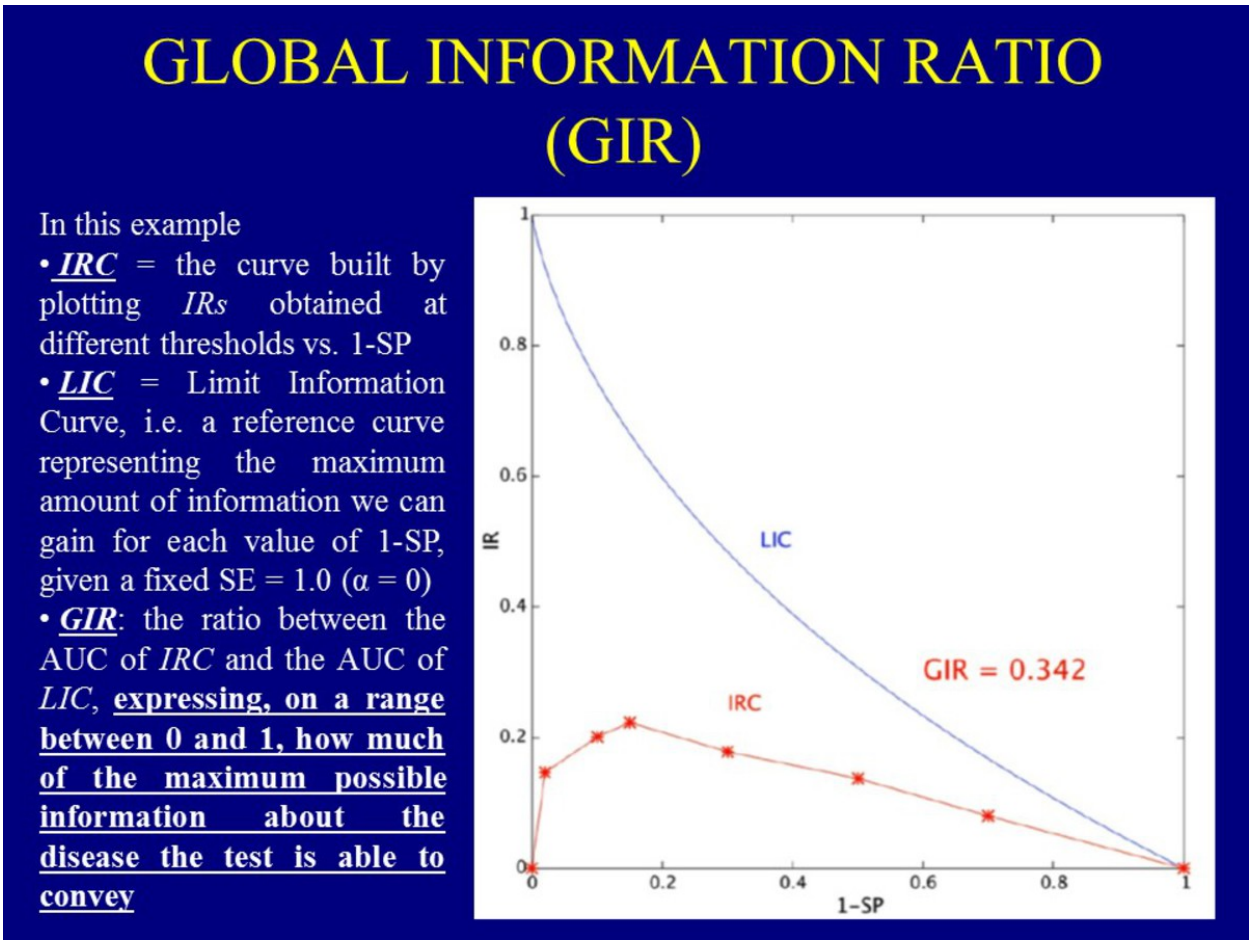


Fig. 9: Properties of GIR.

APPLICATION TO A REAL SCENARIO

	Film Mammography						
Score	7	6	5	4	3	2	1
IR	0.019	0.056	0.093	**0.178**	0.170	0.098	0.000
TP	13	37	62	136	171	204	335
FN	322	298	273	199	164	131	0
TN	42406	42401	42356	41488	39232	32355	0
FP	4	9	54	922	3178	10055	42410
SE	0.039	0.110	0.185	0.406*	0.510	0.609	1.000
1-SP	0.000	0.000	0.001	0.022*	0.075	0.237	1.000

	Digital Mammography						
Score	7	6	5	4	3	2	1
IR	0.015	0.042	0.078	0.178	**0.201**	0.115	0.000
TP	10	28	53	138	187	212	334
FN	324	306	281	196	147	122	0
TN	42235	42224	42180	41204	39029	32466	0
FP	1	12	56	1032	3207	9770	42236
SE	0.030	0.084	0.159	0.413*	0.560	0.635	1.000
1-SP	0.000	0.000	0.001	0.024*	0.076	0.231	1.000

Table 1: The table shows IR, SE and 1-SP values for each cut-off point of the 7-point scale. Calculation was performed based on TP, TN, FP and FN values extracted from the reference work. In the case of DM, the informational cut-off of best accuracy corresponds to additional 49 TPs and 2175 FPs as compared to the ROC cut-off.

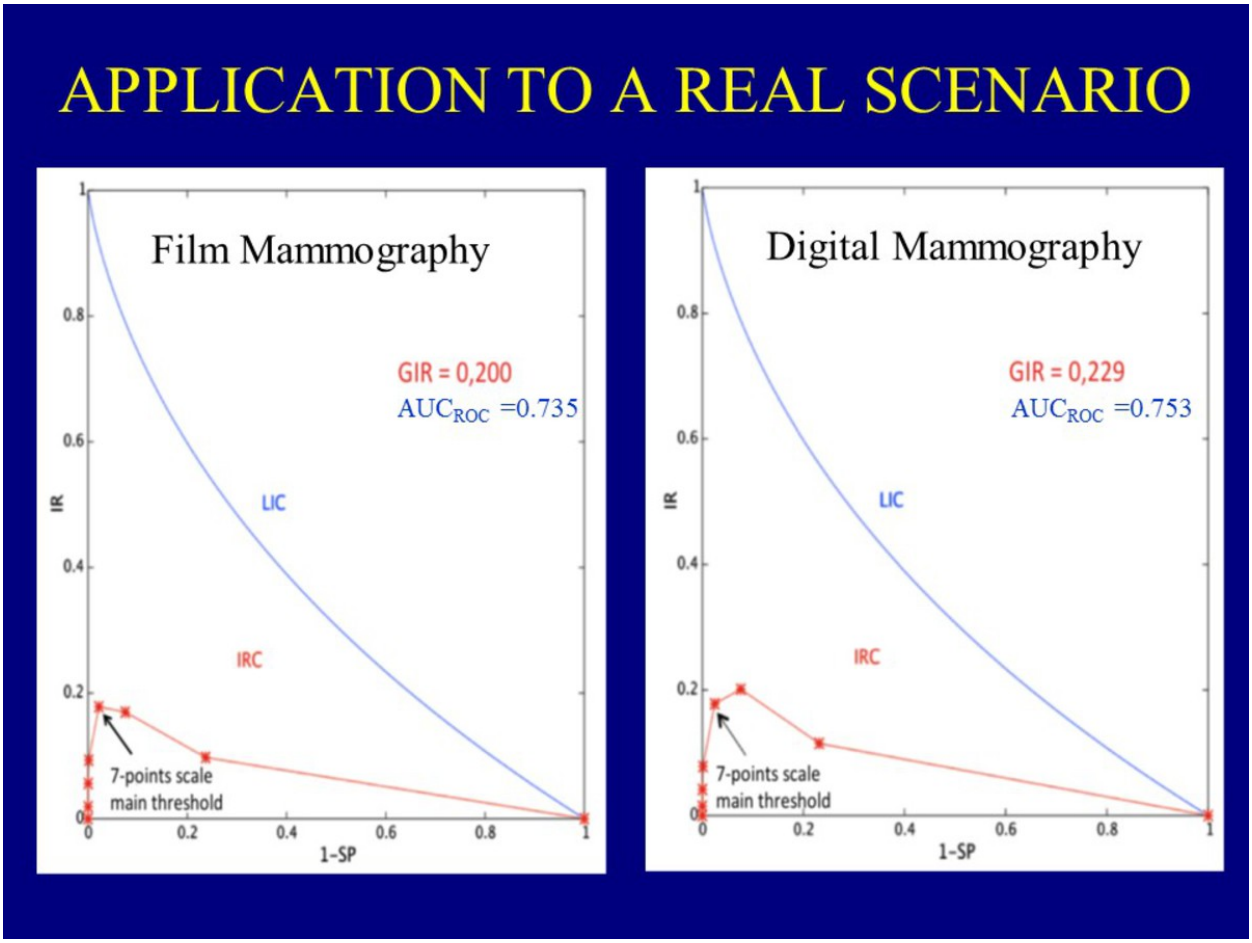


Fig. 10: Comparison between GIR and AUCs of informational and ROC analysis, respectively.

MATHEMATICAL APPENDIX (I)

The bond between D and R is fixed by the transition matrix of the diagnostic channel , and the flow of information between D and R is measured by the *Mutual Information* ; ; it is defined as

$$I(D,R) = \sum_{\substack{d\in\mathcal{D}\\ r\in\mathcal{R}}} p(d,r)\log\frac{p(d,r)}{p(d)p(r)} \tag{1}$$

where $p(d,r) = \Pr\{D=d, R=r\}$ $(d\in\mathcal{D},\ r\in\mathcal{R})$ are the *joint* probabilities, $p(d) = \Pr\{D=d\}$ and $p(r) = \Pr\{R=r\}$ are the *marginals*, and the logarithm is taken to the base 2. By taking into account of the Bayes rule $p(d,r) = p(d)p(r/d)$, and that $p(r) = \sum_{d\in\mathcal{D}} p(d,r) = \sum_{d\in\mathcal{D}} p(d)p(r/d)$ as its consequence, we have

$$I(D,R) = \sum_{\substack{d\in\mathcal{D}\\ r\in\mathcal{R}}} p(d)p(r/d)\log\frac{p(r/d)}{p(r)} = \sum_{\substack{d\in\mathcal{D}\\ r\in\mathcal{R}}} p(d)p(r/d)\log\frac{p(r/d)}{\sum_{d\in\mathcal{D}} p(d)p(r/d)} \tag{2}$$

When decoded in terms of SE , FNR , FPR and SP we have

$$\begin{aligned} I(D,R) = {} & PREV\cdot SE\cdot\log\frac{SE}{PR} + PREV\cdot FNR\cdot\log\frac{FNR}{NR} + \\ & + (1-PREV)\cdot FPR\cdot\log\frac{FPR}{PR} + (1-PREV)\cdot SP\cdot\log\frac{SP}{NR} \end{aligned} \tag{3}$$

where $PREV$ is the pre-test probability of disease, while PR and NR are, respectively, the probability of a positive and of a negative test result.

Fig. 11: Mathematical definition of I(D,R) in diagnostic terms.

MATHEMATICAL APPENDIX (II)

To evaluate the best performance of a diagnostic test, we have to maximize the Mutual Information (6) for the binary asymmetric diagnostic channel , over all possible pre-test probabilities of disease; this leads to the maximum amount of information that a diagnostic test can provide, that is to the capacity of the channel. After some mathematical manipulation , the *MI* can be expressed as

$$I(D,R) = h(u) + (h(\alpha) - h(\beta))p_D(0) - h(\alpha) \quad (4)$$

where $h(.)$ is the binary Shannon entropy function defined as

$$h(p) = -p\log_2 p - (1-p)\log_2(1-p) \quad (5)$$

and $u{=}\alpha{+}p_D(0)(1{-}\alpha{-}\beta)$ $(0{\leq}u{\leq}1,\ \alpha{+}\beta{<}1)$. Note that $I(D,R)$ is in the range $0\ldots1$. We then have

$$C_{BAC} = \frac{\beta}{1-\alpha-\beta}\cdot h(\alpha) - \frac{1-\alpha}{1-\alpha-\beta}\cdot h(\beta) + \log_2(1{+}z) \quad (6)$$

where

$$z = 2^{\frac{h(\beta)-h(\alpha)}{1-\alpha-\beta}} \quad (7)$$

The pre-test probability of disease $p_D^*(1)$ that achieves this capacity is given by

$$p_D^*(1) = \frac{(1-\beta)(1+z)-1}{(1-\alpha-\beta)(1+z)} \quad (8).$$

Fig. 12: Mathematical definition of the «Capacity» of a diagnostic channel.

MATHEMATICAL APPENDIX (III)

This approach requires to evaluate the definite integral of function (4) . If we set $p_D(0) = x = 1 - p_D(1)$ and $MI = I(D, R)$ for sake of simplicity, we need to calculate

$$\int_0^1 MI \, \mathrm{d}x = \int_0^1 (h(u) + (h(\alpha) - h(\beta))x - h(\alpha)) \, \mathrm{d}x \tag{9}$$

where $u = \alpha + x(1 - \alpha - \beta)$ $(0 \leq u \leq 1, \alpha + \beta < 1)$. This integral can be evaluated by parts; here is the final value

$$\int_0^1 MI \, \mathrm{d}x = \frac{1}{1-\alpha-\beta} \left\{ \frac{(-\beta)^2 \log(\beta) + (1-\beta) - (1-\beta)^2 \log(1-\beta)}{\log 4} - \frac{(\alpha-1)^2 \log(1-\alpha) + \alpha - \alpha^2 \log(\alpha)}{\log 4} \right\} - \frac{h(\alpha) + h(\beta)}{2} \tag{10}$$

where the log are now to the natural base e.
Note that the *MI-Curve* area of the standard of reference can be computed by substituting $\alpha = \beta = 0$ in the formula (10); his leads to

$$\int_0^1 MI \, \mathrm{d}x \Big|_{\alpha=\beta=0} = \frac{1}{\log 4} = 0.7213 \tag{11}.$$

Fig. 13: Mathematical definition of the AUC under the MI-curve.

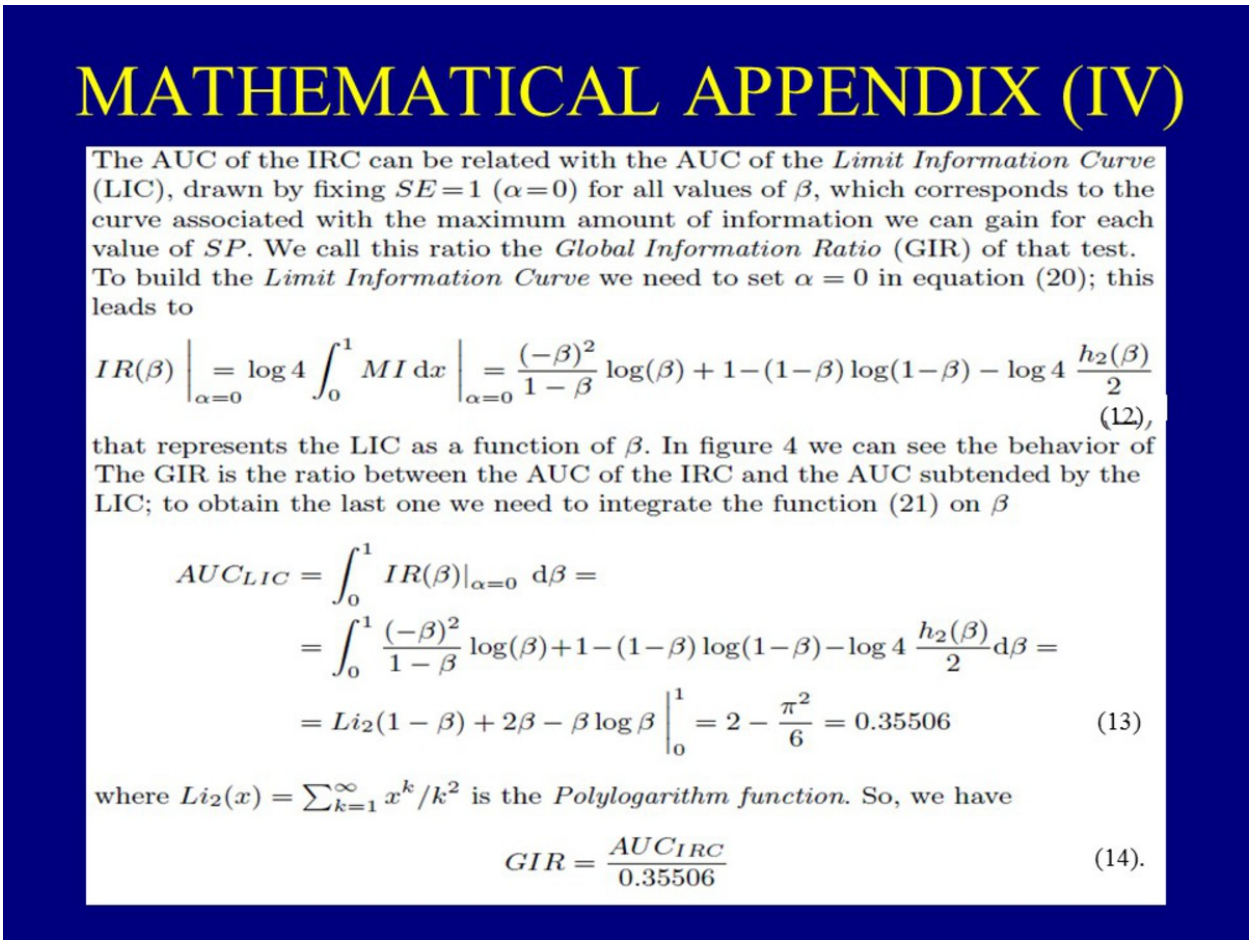


Fig. 14: Mathematical definition of GIR.

Conclusion

1. Using concepts from IT (i.e., the most influencing mathematical theory underlying telecommunication systems), the radiological diagnosis ca be modeled as an "asymmetric binary transmission channel". In this model, information is "transmitted" from the disease to the radiologist through a "diagnostic channel" represented by an imaging tool.

2. Using the above model, the accuracy of a diagnostic test is expressed in informational terms, i.e. with rigorous mathematical measures of information derived from Mutual Information (*MI*).

3. We believe that *informational analysis* might complement or replace standard analysis of accuracy when assessing whether the performance of a diagnostic test:

• has been conditioned by definite characteristics of the clinical setting (low prevalence, selection bias and so on) or rather by intrinsic limits of the amount of diagnostic information that the test can physically convey;

• can be further improved for a certain clinical scenario, or rather mirrors the maximum limit of information that can be obtained given the test characteristics. This might lead to changes in diagnostic strategy.

Personal information

Dr. Rossano Girometti, MD

Researcher,

Institute of Radiology, Department of Medical and Biological Sciences, University of Udine

via Colugna, 50 - 33100 Udine

Italy

rgirometti@sirm.org

Prof. Francesco Fabris, PhD

Associate Pofessor of Informatics,

Department of Mathematics and Earth Sciences, University of Trieste

via Valerio, 12b - 34127 Trieste

Italy

ffabris@units.it

R. Girometti and F. Fabris contributed to the paper equally.

References

1. Weinstein S, Obuchowski NA, Lieber ML. Clinical evaluation of diagnostic tests. Am. J. Roentgenol. 2005;184:14-9.

2. Bewick V, Cheeck L, Ball J. Statistic review 13: Receiver operating characteristic curves. Critical Care 2004;8:508-12.

3. Shannon CE. A mathematical theory of communication. The Bell System Technical Journal, 1948; 27:379-423, 623-56, July.

4. Verdu S. Fifty years of Shannon Theory. IEEE Transactions On Information Theory 1998;44:2057-78.

5. Cover TM, Thomas JA, Elements of Information Theory, John Wiley & Sons, Inc. 1991.

6. MacKay DJC. Information Theory, Inference, and Learning Algorithms. Cambridge: Cambridge University Press, 2003

7. Pisano ED, Gatsonis C, Hendrick E, et al. Diagnostic performance of digital versus film mammography for breast-cancer screening. N Engl J Med 2005;353:1773-1183

8. Obuchowsky NA. ROC analysis. Am. J. Roentgenol¨. 2005;184(2):364-72.