1259 UNICO

COMPENDIUM
OF SCIENTIFIC ARTICLES FROM THE

LATIN AMERICAN CENTER FOR PERINATOLOGY AND HUMAN DEVELOPMENT

PAN AMERICAN HEALTH ORGANIZATION
REGIONAL OFFICE OF THE
WORLD HEALTH ORGANIZATION

EVALUATION OF DIAGNOSTIC PROCEDURES. METHODOLOGICAL ASPECTS

R. FESCINA, F. SIMINI; R. BELITZKY

Medical science has witnessed great technological advances during the last decades affecting Perinatology in particular. New diagnostic procedures have been incorporated as a compliment to those presently in use. Better diagnostic methods have reduced the need for simultaneous tests, which were uncomfortable and risky for the patient as well as producing confusing or contradictory results and increasing medical costs.

Considering the natural human tendency, of believing more in procedures than in criteria, it may be beneficial to remember the methodology to be followed in the evaluation of any diagnostic procedure.

To make a diagnosis means to corroborate the existence of a particular morbid state. In clinical practice, the diagnosis is made through two different phases. During the first phase, a presumption, suspicion or hypothesis of the existence of the disease is esta-blished by ranking clinical signs and symptoms and their logical association with known pathologies. The second phase is directed toward this presumption, suspicion or hypothesis. With this purpose in mind, discernment is done by:

a) Tests and examinations that, if positive, confirm the presence of illness, and

b) Tests and examinations that, if negative, rule out the presence of the illnesses and of similar symptoms.

Due to uses and practices, there is general consensus in associating the positive results of a test with the presence of a disease, a negative result with the absence of disease or illness

Clinical information should constitute the basic pre-requirement in the orientation of what tests should be indicated and in the interpretation of results. Therefore, additional information provides more security and the clinical knowledge increases, but it is not a substitute for further diagnostic procedures.

Development of a Diagnostic Procedure

In 1947, Yerushalmy was surprised by the distinct and antagonistic diagnoses made by very experienced radiologists when studying the same lung X-rays, and noted that there was no way to measure the validity or error in these X-rays. With this idea in mind, he recorded the diagnoses, and made a follow-up of cases, until he confirmed the presence or absence of the illness by anatomical pathology obtained in surgical procedures or in autopsies. Using this methodology, he established outlines for evaluating the degree of confidence of the procedures or diagnostic tests. When doing the tests or exams, 2 questions arose:

If illness is present, what is the probability of the result being positive?, and

In the absence of illness, what is the probability of the results being negative?

The answer to the first question, determines ***sensitivity***, and the second question, ***specificity***.

Sensitivity, is the procedure's capability to give a correct diagnosis of the disease when it is present (True Positive or illness).

Specificity, is the procedure's capability to make a correct diagnosis in the absence of illness, (True Negatives or without a specific pathology).

As these conclusions are qualitative, since they indicate the confirmed presence or absence of a disease, and/or the Positive or Negative results of a diagnostic procedure, double entry tables are used (decision matrix) for its numerical presentation *(Figure 1).*

a) Cases with positive diagnosis in presence of disease *(True Positive).*

b) Cases with a positive diagnosis in absence of the disease *(False Positive).*

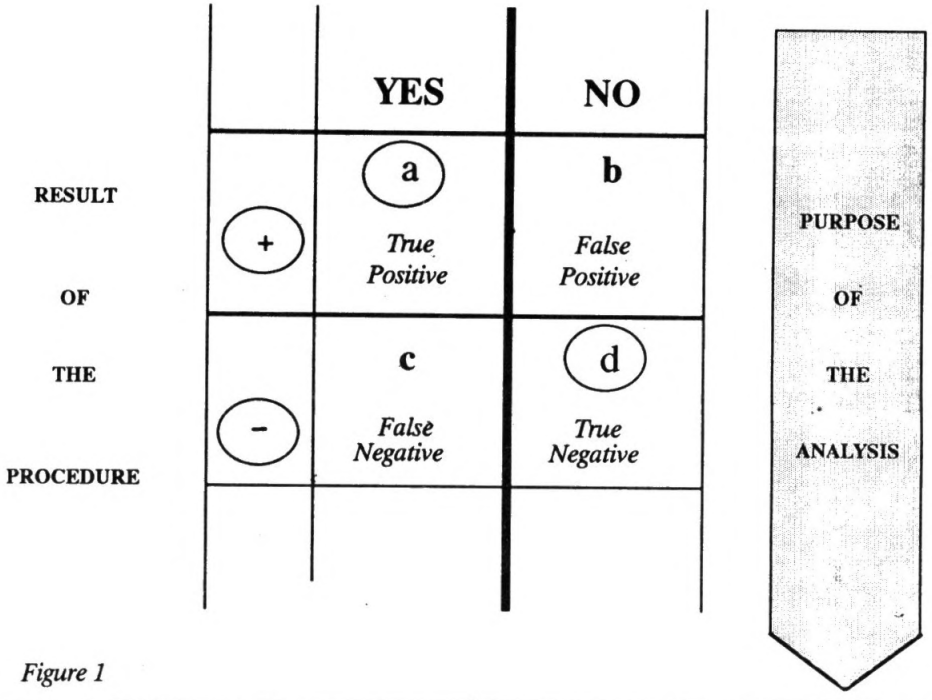


Figure 1

c) Cases with a negative diagnosis in presence of disease *(False Negatives)*.

d) Cases with negative diagnosis in the absence of disease *(True Negatives)*

Sensitivity is measured by the proportion of individuals whose results of the procedure were Positive and their relation to the total number of individuals who have the disease. It is the analysis of the first column of the table (vertical position).

$$\text{Sensitivity} = \frac{\text{True Positives}}{\text{Total Number of Ill Individuals}} \times 100 = \frac{a}{a + c} \times 100$$

Specificity is measured by the proportion of individuals with a Negative diagnosis and without pathology in relation to the total who do not have the disease (analysis of the second column of the table).

$$\text{Specificity} = \frac{\text{True Negatives}}{\text{Total without pathology}} \times 100 = \frac{d}{b + d} \times 100$$

Sensitivity covers *True Positive* while specificity considers *True Negatives*

The same decision matrix allows for an evaluation of errors, one way or the other. Therefore, from the test's point of view, these are called false negatives, when the procedure's results are negative .

but the disease is present. This type of error is measured through the following:

$$\text{False Negative Rate} = \frac{\text{False Negatives}}{\text{Total Number of Ill Individuals}} \times 100 = \frac{c}{a+c} \times 100$$

In other words this rate constitutes the complement to sensitivity to reach 100%. Therefore, it can be established as follows:

100 - Sensitivity = False Negative Rate

***False Positives** are* those results which are positive, but there is an absence of disease. This type of error is measured through the following method:

$$\text{False Positive Rate} = \frac{\text{False Positives}}{\text{Total without pathology}} \times 100 = \frac{b}{b+d} \times 100$$

This rate represents the complement of 100% specificity, therefore, it can be estimated as follows:

100% - Specificity = False Positive Rate

The closer to 100% sensitivity and 100% specificity, the higher the test's capability to discern those who are ill from those that are not ill. Except those cases which are very relevant examples (death diagnosis through prolonged register of *EEG*), the existence of error in either way, must be considered.

Figure 2 illustrates an ideal test in which distributions with and without a pathology do not superimpose their values. This event rarely occurs for the majority of tests.

For example, if the interest is to determine the level of postprandial blood glucose that will differentiate pregnant diabetic mothers from those who are not diabetics, it is possible to calculate specificity and sensitivity for different cutting points, according to the investigator's convenience.

As seen in *Figure 3,* choosing level 1, specificity is closer to 100% (the non diabetics are correctly separated but the sensitivity is low because a certain

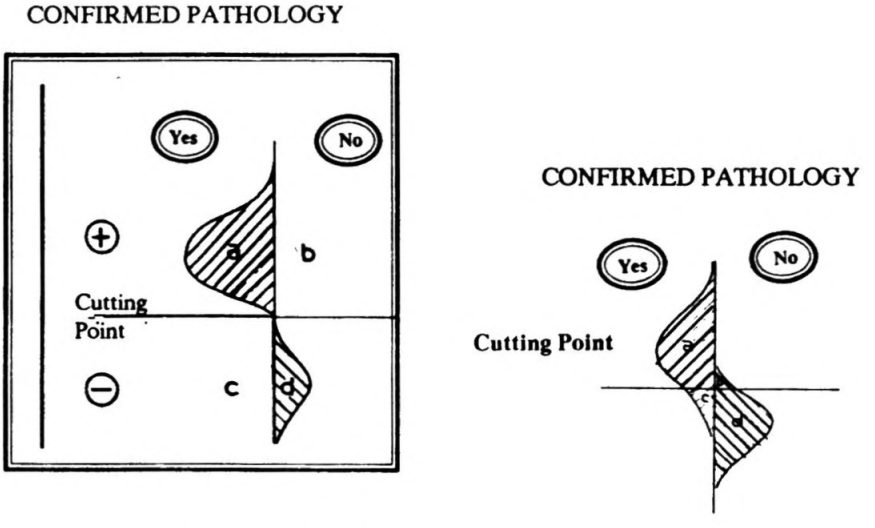
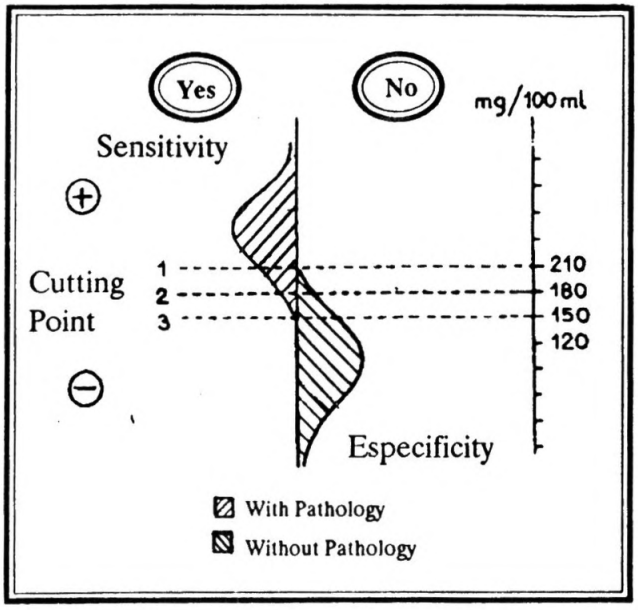


Figure 2 : Ideal Test: Sensitibity and Specificity = 100% (only correct cases of present or absent pathology)
Majority of the Tests: Sensitibity and Specificity less than 100%. There are diagnosis errors (false positive and false negatives)

Figure 3 - **POSTPRANDIAL GLYCEMIA**



proportion escape the diagnosis (there are *false negatives).*

Choosing level 3, sensitivity is close to 100% (practically all women with diabetes were detected, but the non diabetics are not well discriminated *(low specificity).*

At this stage, the investigator that knows who is diabetic, determines to his own convenience the sensitivity and specificity, selecting the cutting point.

As sensitivity is increased, specificity is lost and vice-versa, therefore, both values must be examined together

To develop a diagnostic procedure, it is indispensable to have a final confirmation, without doubt, about the presence or absence of pathology. This allows us to calculate the sensitivity and specificity. The original investigation is directed mainly to use the procedure in studying individuals whose correct diagnosis can finally be known.

Sensitivity and specificity do not vary with the prevalence of illness, but they do with the cutting point

Practically, diagnostic procedures are used in a very different manner than for the purpose for which they are developed. Physicians use these procedures in patients with unknown diagnosis, and the purpose of performing these tests is to help determine the probable condition of the patient. Work is done under conditions of uncertainty, and these tests are used to predict the presence or absence of a disease, so there is always a margin for error. In order to affirm the presence of an illness or disease, through a diagnostic procedure, the following facts have to be taken into consideration:

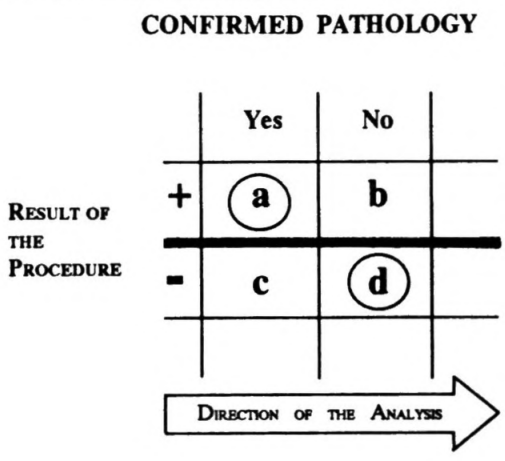
a) If the procedure's results are *Positive* what are the probabilities that a disease is really present? (that the illness is confirmed)?

b) If the results are *Negative,* what is the probability that illness is not really present? (that the disease is ruled out)?

The same decision matrix already prepared, provides the answer to both questions, but in this case, the analysis is done by rows. (horizontally).

Predictive Value of the Positive Test (P.V.P.T.)

The probability that illness is present when the procedure's result is Positive (analysis of the first horizontal row) is expressed as the percentage of truly ill individuals among those that had positive tests.



$$P.V.P.T. = \frac{\text{True Positive}}{\text{Total Positive Tests}} \times 100 = \frac{a}{a+b} \times 100$$

Predictive Value of the Negative Test (P.V.N.T.)

Probability dictates that illness is not present when the result of the procedure is negative (analysis of the second row). This is expressed as a percentage of patients without the pathology in relation to all the patients that had negative tests.

$$P.V.N.T. = \frac{\text{True Negatives}}{\text{Total of Negative Tests}} \times 100 = \frac{d}{c+d} \times 100$$

Error Measurement

It is observed that there are cases in which the test indicates the presence of a disease, and there are other cases in which it does not, ***(false positive*** in the prediction) and cases where the results of test are negative, but the disease is truly present, ***(prediction's false negatives)***. These errors are measured in the relations:

False Positive of a Positive Test (F.P.P.T.)

$$F.P.P.T. = \frac{\text{False Positive}}{\text{Total of Positive Tests}} \times 100 = \frac{b}{a+b} \times 100$$

It is the complement to 100% of the Predictive Value of the Positive Test (P.V.P.T.).

False Negatives of the Negative Test (F.N.N.T.)

False Negatives to the *Negative Test* are the individuals that are sick but were classified by the procedure as non-ill (c) (analysis of the second horizontal row) in relation to the total negative tests.

$$F.N.N.T. = \frac{\text{False Negative}}{\text{Total of Negative Tests}} \times 100 = \frac{c}{c+d} \times 100$$

It is the complement of 100% of the Predictive Value to the Negative Test (P.V.N.T.).

Also, accuracy is used. It is a rate that jointly considers the correct predictions of the procedure, positive and negative.

$$Accuracy = \frac{\text{True Positives + True Negatives}}{\text{Total individuals with and without the pathology}} \times 100 = \frac{a+d}{a+b+c+d} \times 100$$

This ratio allows a comparison of different procedures, but its disadvantage is that it grants *False Positives* and *False Negatives,* the same value, and these errors may have different repercussions in clinical practice.

It is important to emphasize that the ***predictive value*** of a test and its *accuracy* are substantially modified when the *prevalence varies.*

A high predictive value can be expected with a high prevalence (e.g. epidemics), with ***low prevalence*** prediction decreases, even though sensitivity and specificity do not change.

The following example *(Figure 5)* shows how prevalence changes the predictive value of a test (this particular test is characterized by a given sensitivity and specificity).

CONFIRMED PATHOLOGY

	YES	NO	
+	40	2	42
-	10	23	33
	50	25	75

$$\text{Prevalence} = \frac{50}{75} \times 100 = 67\%$$

$$\text{Sensitivity} = \frac{40}{50} \times 100 = 80\%$$

$$\text{Specificity} = \frac{23}{25} \times 100 = 92\%$$

$$\text{P.V.P.T.} = \frac{40}{42} \times 100 = 95\%$$

$$\text{P.V.N.T.} = \frac{23}{33} \times 100 = 70\%$$

CONFIRMED PATHOLOGY

	YES	*NO*	
+	*40*	*20*	*60*
-	*10*	*230*	*240*
	50	*250*	*300*

$$\textit{Prevalence} = \frac{50}{300} \times 100 = 17\%$$

$$\textit{Sensitivity} = \frac{40}{50} \times 100 = 80\%$$

$$\textit{Specificity} = \frac{23}{25} \times 100 = 92\%$$

$$\textit{P.V.P.T.} = \frac{40}{60} \times 100 = 67\%$$

$$\textit{P.V.N.T.} = \frac{230}{290} \times 100 = 96\%$$

Prevalence corresponds to the number of ill individuals in relation to the total number of "cases" considered.

It is observed that when there is a decrease in the prevalence, the P.V.P.T. decreases and the P.V.N.T. increases. This clearly indicates that one cannot directly compare the predictive values of the developed procedures in different prevalence groups. *Bayes' Theorem* of conditioned probability permits one to adjust the predictive value of a given prevalence of an illness.

$$P.V.P.T. = \frac{\textit{Prevalence of Illness x Sensitivity}}{\textit{Prevalence x Sensitivity + (1-Prevalence)(1-Specificity)}}$$

SECOND STAGE:

PROCEDURE PROOF

Once the diagnostic procedure has been developed

in its theoretical phase, it must be tested prospectively. It is important to consider the groups with different prevalence of the different diseases and to verify if results are reproducible by other research groups in other places. This way advantages and limitations can be known, as well as studying reliability and precision under different circumstances (medical aid complexity, technical personnel quality, different equipment, etc.).

It is useful for the procedure to be replicated by the same group that developed it, and that there be exchange of information between the research groups. A dissemination and extensive use of the procedure is not advisable, until it has satisfactorily passed this evaluation phase.

THIRD STAGE: *UTILIZATION OF THE PROCEDURE*

With the application of a new diagnostic procedure, an effort is being made to improve the results attributable to mortality and morbidity. The principal requirement needed is the capability to change the final results: A correct and timely diagnosis leads to a more successful intervention.

To evaluate procedures factors that differ from those used in previous stages, certain elements must be considered: Infrastructural characteristics, availability of personnel and equipment that would allow their feasibility; financial and human costs (aggressiveness, pain, etc.) and ethical aspects regarding the distribution of benefits such as is treatment available for all who need it or only for those who can pay for it?.

At this evaluation stage, these questions must be answered:

- *Is the developed procedure better than the previous ones?*
- *It is easier, faster, simpler, etc*
- *Does it improve patient management and treatment?*
- *Treatment is more convenient, more specific, less painful, etc.*
- *Does it permit a faster recovery?*
- *Fewer days of illness, less cost, etc.*
- *Does it contribute to improving the overall health of the population?*

PURPOSE OF THE PROCEDURES OR DIAGNOSTIC TESTS

These tests are specifically used for three purposes: A) To detect the disease; B) To confirm the disease; C) To rule it out.

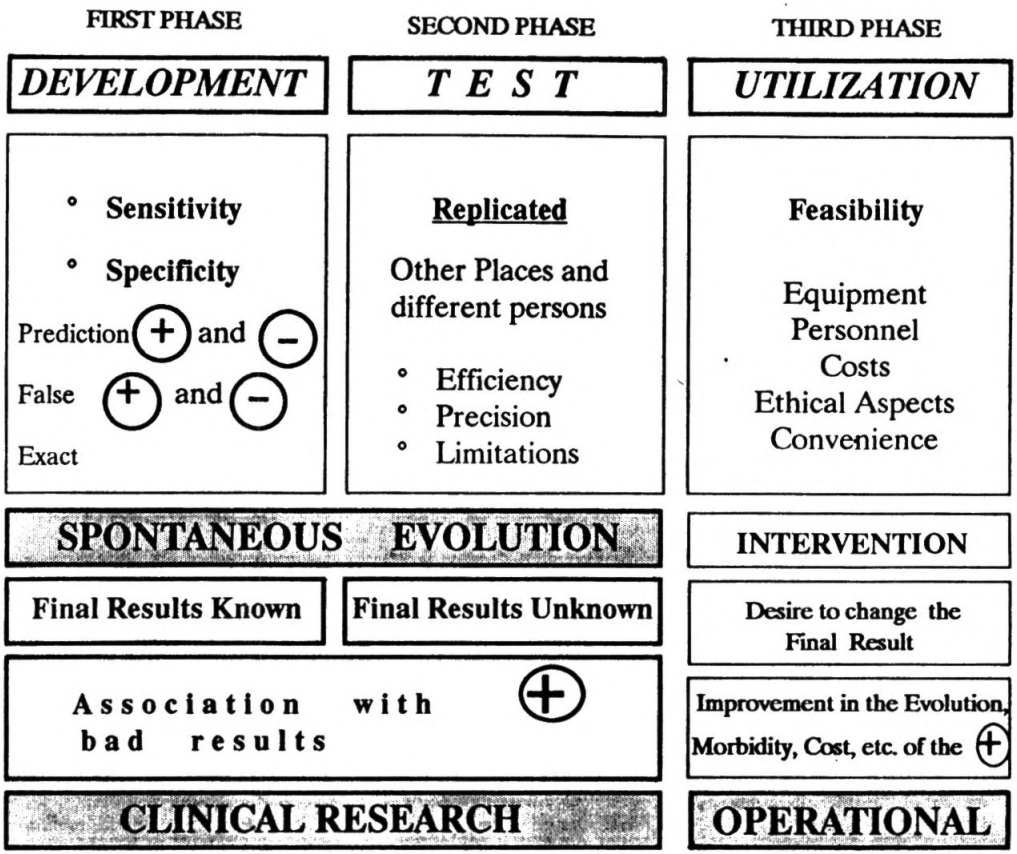
A) **Disease Detection Tests** *(SCREENING)*

An effort is made to apply the diagnostic procedure in an apparently healthy population (those who do not present any symptoms or any evidence of clinical signs of the illness) to detect cases at their early stages. Some known examples are the testing for fasting blood sugar to detect diabetes, or the Papanicolaou smear to detect genital cancer. **High sensitivity** tests are required for screening tests (that would detect the highest possible number of ill individuals) which are easy to do, quicker, non invasive and of low cost. A certain number of errors or **false positives** are accepted in the test. Thus, the **screening procedures**, when positive, **always require other tests to confirm the disease.**

The cases with a positive results must not be directly considered as having the disease. Assessments must be carried out when confirmatory tests are required, and these must range from the most simple to the most complicated.

B) *Tests to Confirm the Illness*

These are used in situations when there is strong suspicion of illness. Their purpose is to verify this suspicion. These procedures are usually more complicated and aggressive, such as the bronchoscopy, or the cervical-uterine biopsy, if a neoplasm is suspected. These tests must be **highly specific** so there are practically no false positives. High positive predictive value is required. In general, these procedures are costly and are done by specialists, when there is a certain diagnostic presumption and when a further confirmation is desirable in order to perform risky surgical procedures or aggressive treatments.



This methodology is not restricted necessarily to instrumental or laboratory tests, and can be used for any criteria that may be useful in the identification of pathological conditions. If one pretends to predict or to make a prognosis of the final result of birth, for example (depression at birth, morbidity or death, etc.), it is essential to abstain from interventions that can modify the result during the first and second stage. This methodology can also be applied to calculate rates or classifications with the objective of instituting different treatments according to risk.

THE PROCESS TO RULE OUT DIAGNOSES

Directed to	Appropriate to do:
Asymptomatic People who is not ill or at a stage of illness in a pre-clinical or incipient period	Detection tests, of reasonably high sensitivity, easy, non-aggressive and at low cost. Many false positives are accepted. The detected cases must be confirmed.
Symptomatic patients Patients with Illness and symptoms corresponding to a determined morbid status. These are found by suspicious pathology.	Confirmation of specificity tests close to 100% to avoid false positives. Positive results indicate illness. It is acceptable that test results originate some concern, are expensive and that some risk is implied.
Not ill and symptomatic Symptomatic and with questionable tests. Patients with Similar Illnesses. The probability of existence of pathology is not very high, but the importance of this favors its dismissal.	Exclusion Test with sensitivity near 100% to avoid false negatives. The negative cases are considered as patients that are not ill. There may be concerns about complications.

C) ***Tests to Rule Out Illness***

The tests used are those in which the Negative results would rule out the possibility of the suspected illness. Very high sensitivity (almost 100%) is required, with no false negatives in the test.

The majority of these procedures are expensive, sophisticated or annoying, for example, colon examination or rectosigmoidoscopy to rule out the presence of colon cancer in patients with rectal bleeding.

COMBINATION OF TESTS

In practice, more than one procedure is often used. however, this depends on each particular case and on the availability of tests. In general, the diagnostic procedure is done by stages. In accordance with results obtained, other procedures may be indicated. This is called *series* or *tandem* testing.

It is infrequent that a single diagnostic test would accomplish the detection and confirmation purposes. A high sensitivity test is used to detect possible patients and a follow-up is made on cases with positive results through high specificity tests, thus avoiding false positives. Results are due to good judgement and the best timely sequence of tests.

In general, when the serial diagnostic tests are used for detection purposes, specificity is gained, and when parallel technique is used (all tests to all cases) sensitivity is gained.

Currently there is a wide proliferation of tests. It is therefore recommended that for the routine utilization of a procedure, it's specificity, sensitivity and predictive value characteristics be studied. Thus, it is important to know the original tests performed and the characteristics of the population studied. It is also useful to establish certain standardization for the logical application of a *battery (group) of tests* according to the situations presented, whether it be detection of the pathology in patients who do not seem ill, or to confirm or exclude the illness when there is a suspicion that it does exist. In these cases, less aggressive procedures for the patient should be selected, and the information obtained from these procedures should render a higher degree of security. These selections must not be personal decisions, but they must be predicted and standardized by the health team. This practice contributes to the validity of the evaluation.

For a better interpretation of the use of this methodology an exercise is presented.

A group of 163 pregnant women with amenorrhea and a single fetus where studied, and the uterine height was measured weekly.

The diagnosis of Hypogrowth (hypergrowth) was done when the evolution of the uterine height showed a low profile (decrease on the rate of growth with one or more values under the chosen percentile). Newborns were classified by their neonatal weight curve, and by the gestational age established by C.L.A.P. The classification of retarded growth was obtained when the neonate, weighed less than the established value of the 10th percentile corresponding to its age.

Of the 163 babies born, 42 had a confirmed diagnosis of intra-uterine growth delay (IGD) and 121 had adequate weight for their age. When choosing the discriminating point of P10 of uterine height, decision matrix A is constructed. If P25 is chosen from the same uterine height curve, matrix B is obtained.

INTRAUTERINE GROWTH DELAY

A	Yes	No	
Less than P10	*22*	*10*	*32*
Greater or equal than P10	*20*	*111*	*131*
Total:	*42*	*121*	*163*

Sensitivity = *P.V.T. Positive =*
Specificity = *P.V.T. Negative =*
Prevalence =

INTRAUTERINE GROWTH DELAY

B	Yes	No	
Less than P25	*29*	*27*	*56*
Greater or equal than P25	*13*	*94*	*107*
Total:	*42*	*121*	*163*

Sensitivity = *P.V.T. Positive =*
Specificity = *P.V.T. Negative =*
Prevalence =

Estimate, compare, decipher, and make a decision of which of the two uterine high percentiles (P10 or P25) should be chosen. The solution is presented as follows:

EFFICIENCY OF INTRAUTERINE HEIGHT (UH) IN THE INTRAUTERINE GROWTH DELAY

	A Considering the P10 of UH	B Considering the P25
Sensitivity	*52%*	*69%*
Specificity	*92%*	*78%*
P.V.T. Positive	*69%*	*52%*
P.V.T. Negative	*85%*	*88%*
Prevalence of the IGD	*25.8%*	*25.8%*

With P10 of Uterine Height (U.T.) for age determination, one of each two cases with intrauterine delayed growth can be diagnosed (52% sensitivity) and affirm, with an 8% error (92% specificity), that the fetus is not small for its gestational age. The uterine height method is excellent to separate the unaffected group by its height specificity. The predictive value of the positive test for a prevalence of IRG of 25.8% is 69%.

Considering as a discriminating limit the uterine height percentile of 25, sensitivity is improved (69%) with a decrease of the specificity and of the positive predictive value.

BIBLIOGRAPHY

1. BENNET, B.M. "On comparisons of sensitivity, specificity and predictive value of a number of diagnostic procedures". **Biometrics**. 28:793-800, 1972.

2. BELITZKY, R. "Desarrollo, prueba y evaluación de los Procedimientos diagnósticos." Guía de discusión. Montevideo, CLAP, 1983. (Publicación Científica CLAP, 995)

3. FEINSTEIN, A.R. "Clinical Biostatistics." Mosby, Saint Louis, 1977 pp. 214-226.

4. FESCINA, R.H., BELITZKY, R., SIMINI, F. "Evaluación de los Procedimientos diagnósticos." Montevideo, CLAP-OPS/OMS, 1983. (Publicación Científica CLAP 999).

5. MANTEL, N. "Evaluation of a class of diagnostic test." Biometrics 7:240-246, 1951.

6. SIMINI, F. "Eficacia de los procedimientos diagnósticos." Curso de Metodología Científica del CLAP, Montevideo, CLAP-OPS/OMS, 1979.

7. VECCHIO, T.J. "Predictive value of a single diagnostic test in unselected populations." New England Journal of Medicine 274:1171-1173, 1966.

8. YERUSHALMY, J. "Statistical problems in assessing methods of medical diagnosis, with special reference to X-ray techniques." **Pub. Health. Rep.** 62:1432-1449, 1947.