Methodology for estimation of disease prevalence and sensitivity and specificity of a diagnostic test without a standard reference test.

Georgi Gachev
Institute of Mathematics and Informatics - Bulgarian Academy of Sciences, Sofia, Bulgaria

ABSTRACT:
The COVID-19 pandemic raised a question about clinical accuracy of diagnostic tests. This article presents a method for estimating the important measures of a diagnostic test without a standard reference. The sensitivity and specificity along with prevalence disease are estimated based on experimental probability and latent class model analysis. A mathematical model and a computer simulation are developed to test and validate the proposed methodology. An example of implementation and evaluation is given.

Keywords: diagnostic test, prevalence, sensitivity, selectivity, latent class model, example of implementation

INTRODUCTION

The COVID-19 pandemic raised a question about clinical accuracy of diagnostic tests. Estimation of sensitivity *Se* and specificity *Sp* of a diagnostic test is necessary to assess disease prevalence rate *DP*. The task becomes difficult when all these parameters are unknown and the reference standard is not available.

The traditional method to evaluate the performance of a diagnostic test is to establish a group of individuals with known disease status (i.e., classify the individuals as positive or negative by a standard reference method). The sensitivity of the test (true positive fraction) is then calculated using the diseased individuals, and the specificity of the test (true negative fraction) using the individuals classified by the standard reference test as disease-free. The problem with this approach is that the availability of a standard reference method with perfect sensitivity and specificity is somewhat questionable for a large group of diseases [1]. However, several other approaches have been developed for evaluation of tests in absence of a reference standard, see [2][3] [4] for a review of existing methods.

The methodology discussed below uses latent class model and simulated experimental observation of the events frequencies. The class of models where the disease status of the individuals is unknown are traditionally referred to as latent class models as the disease status is latent: existing but not presently evident or realized [1]. The first stage of the methodology is essentially a computer simulation of this latent phase of the spread of a disease where random persons in the population are marked as having that disease. The second stage simulates an experimental examination of a population group with a diagnostic test for which sensitivity and selectivity are not accurately determined. The third stage involves finding a theoretical frequency of examination results that is as close as possible to the experimental ones in order to find the values of the unknown disease prevalence and sensitivity and specificity of the diagnostic test.

MATERIALS AND METHODS

A diagnostic test has two basic characteristics sensitivity and specificity. Sensitivity is the proportion of true positives tests out of all patients with a condition [5]. In other words, it is the ability of a test or instrument to yield a positive result for a subject that has that disease. The equation for sensitivity *Se* is [6]:

$$Se = \frac{TP}{TP + FN} \tag{1}$$

where *TP* – true positives and *FN* – false negatives.

Specificity is the proportion of true negatives out of all subjects who do not have a disease or condition [5]. In other words, it is the ability of the test or instrument to obtain negative results for a person who does not have a disease. The formula to determine specificity *Sp* is as follows [7]:

$$Sp = \frac{TN}{TN + FP} \tag{2}$$

where *TN* – true negatives and *FP* – false positives.

Disease prevalence *DP*, sometimes referred to as prevalence rate, is the proportion of persons in a population who have a particular disease or attribute at a specified point in time or over a specified period of time [7]. The formula to determine prevalence is:

$$DP = \frac{TP + FN}{TP + FN + TN + FP} \tag{3}$$

The prevalence *DP* could be interpreted as the unconditional probability of being diseased *D*, *Se* – as the conditional probability to test positive being diseased and *Sp* – as the conditional probability to test negative being healthy or not diseased *ND*.

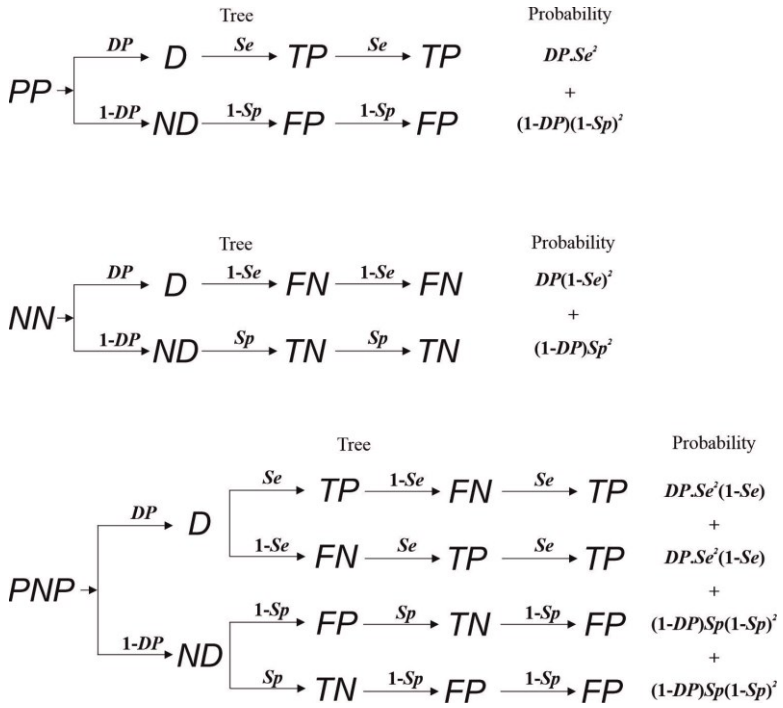
The three unknown variables *Se*, *Sp* and *DP* can be estimated by matching the values of observed experimental probabilities with the corresponding theoretical probabilities. Theoretical probabilities of events that happen in sequence during the diagnostic testing process can be derived from the law of total probability or directly form corresponding parts of the event tree (fig.1.). As the unknown variables are three, it is necessary to choose three independent equations for calculating the event probabilities.

The further step that should be taken is to determine an appropriate sample size group of subjects which will be tested. Depending on the aimed accuracy of the results *n* random subjects will be chosen among the population target group. Each subject will be tested two or three times. The record of consecutive results for each subject will be kept. For the following example tracking of these three consecutive events will be used:

- Twice positive test *PP* – the individuals who have two consecutive positive tests
- Twice negative test *NN* – the individuals who have two consecutive negative tests
- Two alternate tests followed by a positive one *PNP* or *NPP* further both denoted as *PNP* – the individuals who tested twice with different results are tested third time and the test is positive.

The tree of observed diagnostic test results yields the theoretical probabilities of the aforementioned events.
:

Fig. 1. Probability tree for the *PP*, *NN* and *PNP* events



The equations form a system:

$$\begin{cases} PP = DP.Se^2 + (1-DP)(1-Sp)^2 \\ NN = Dp(1-Se)^2 + (1-DP)Sp^2 \\ PNP = 2(Dp.Se^2(1-Se) + (1-DP)Sp(1-Sp)^2) \end{cases} \quad (4)$$

The experimental values of *PP*, *NN* and *PNP* events found in practice or by computer simulation are substituted in (4). The unknown values of Se, Sp, and DP are estimated by solving the system (4) with appropriate tolerance. The roots of interest are in the interval [0,1] as by definition *Se*, *Sp*, *DP* ∈[0,1]. In this interval the system has two roots (*Se*,*Sp*,*DP*) and (*1-Se*,*1-Sp*,*1-DP*).

RESULTS

To check and validate the mathematical model and computer simulation, the calculated values of *Se*, *Sp* and *DP* should be matched to the respective values of known disease prevalence and diagnostic test parameters.

For the following example is supposed that actual disease prevalence is *DP* = 5%, the sensitivity and specificity of an unknown medical test are *Se* = 85% and *Sp* = 95%.

A sample size of $n = 10^4$ individuals are being tested.

The result of the simulation is:

Table 1. Statistical data of the consecutive diagnostic test events *PP*, *NN* and *PNP* of $n = 10^4$ individuals where *DP* = 5%, *Se* = 85%, *Sp* = 95%

Test results (CI = 95%)	Number of individuals			%
	Lower bound	Mean	Upper bound	Relative error
PP – twice positive	349	385	427	10.6
NN – twice negative	8518	8585	8656	0.8
PNP – alternate followed by a positive	133	154	184	17.6

The 876 individuals of the mean value who have been tested as *PNN* or *NPN* are not considered as these events are not included in (4). The probability of each event obtained in Table 1 is substituted in system (4). This system of equations is nonlinear and it is easier to be solved numerically. The solutions of the system on the mean and boundary values of the confidence interval are:

Table 2. Estimated disease prevalence, sensitivity and selectivity calculated from *PP*, *NN* and *PNP* events

Experimental probabilities of the events	%		
	Se	Sp	DP
Lower bound, PP = 0.0349, NN = 0.8518, PNP = 0.0133	88.7	94.2	4
Mean, PP = 0.0385, NN = 0.8585, PNP = 0.0154	**84.9**	**95**	**5**
Upper bound, PP = 0.0427, NN = 0.8656, PNP = 0.0184	81.2	96	6.2
Average relative error	4.4	0.9	22

As could be seen the result for the mean is close to the actual values of the parameters. The deviation of the boundary solutions is due to the relative error of rare events. In the above sample simulation, the *PNP* is less frequent than both other events *PP* and *NN*. The following rare events can occur in sampling phase when trying to estimate a highly accurate diagnostic test:

- P events when the disease prevalence is low
- N events when the disease prevalence is high
- Alternate events
- Any of the events due to small sample size

In every particular case a rare event can be replaced by supposedly more frequent event. Eventually this can be combined with the increase of sample size.

Let's look at previous example and try to replace the *PNP* event in (4) with *PPP* event which is twice as frequent. The equation for the *PPP* event is

$$PPP = DP.Se^3 + (1 - DP)(1 - Se)^3 \quad (5)$$

The simulated process will give:

Table 3. Statistical data of the consecutive diagnostic test event *PPP* with three positive results of $n = 10^4$ individuals where *DP* = 5%, *Se* = 85%, *Sp* = 95%

Test results (CI = 95%)	Number of individuals			%
	Lower bound	Mean	Upper bound	Relative error
PPP – triple positive	273	308	342	11.2

Table 4. Estimated disease prevalence, sensitivity and selectivity calculated from *PP*, *NN* and *PPP* events

Experimental probabilities of the events	%		
	x	y	z
Lower bound, PP = 0.0349, NN = 0.8518, PPP = 0.0273	85.3	94.3	4.4
Mean, PP = 0.0385, NN = 0.8585, PPP = 0.0308	**85.2**	**95**	**5**
Upper bound, PP = 0.0427, NN = 0.8656, PPP = 0.042	84	95.8	5.8
Average relative error	0.8	0.8	14

The average combined deviation of all three parameters in Table 4 is less almost two times compared to the deviation of the same parameters in Table 2.

DISCUSSION:

When trying to apply this method in practice, the following assumptions should be considered:

1. Disease prevalence, selectivity and specificity remain unchanged during the testing period. Supposedly, the selectivity and specificity are constants for a given diagnostic test implemented in similar conditions, which cannot be stated for disease prevalence. The prevalence is variable and changes over the time. For the oncology diseases the changes are slow but for some infectious diseases the changes are fast, so the duration of the testing period should be assessed according to the specific situation.
2. The demonstrated methodology is appropriate when mass diagnostic testing is affordable as in COVID-19 pandemic period.
3. The methodology could be implemented when no other reliable standard test reference is available.

REFERENCES:

1. Toft N, Jørgensen E, Højsgaard S. Diagnosing diagnostic tests: evaluating the assumptions underlying the estimation of sensitivity and specificity in the absence of a gold standard. *Preventive Veterinary Medicine*. 2005 Apr;68(1):19-33. doi: 10.1016/j.prevetmed.2005.01.006. [PubMed - https://pubmed.ncbi.nlm.nih.gov/15795013/]
2. Lewis F, Torgerson P. A tutorial in estimating the prevalence of disease in humans and animals in the absence of a gold standard diagnostic. *Emerging Themes in Epidemiology*. 2012 Dec 28;9(1):9. doi: 10.1186/1742-7622-9-9. [PubMed - https://pubmed.ncbi.nlm.nih.gov/23270542/]
3. Habibzadeh F. On determining the sensitivity and specificity of a new diagnostic test through comparing its results against a non-gold-standard test. *Biochem Med (Zagreb)*. 2023 Feb 15;33(1):010101. doi: 10.11613/BM.2023.010101. [PubMed - https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9807233/]
4 Keddie SH, Baerenbold O, Keogh RH, Bradley J. Estimating sensitivity and specificity of diagnostic tests using latent class models that account for conditional dependence between tests: a simulation study. *BMC Med Res Methodol*. 2023 Mar 10;23(1):58. doi: 10.1186/s12874-023-01873-0. [PubMed - https://pubmed.ncbi.nlm.nih.gov/36894883/]
5. Bolin E, Lam W. A review of sensitivity, specificity, and likelihood ratios: evaluating the utility of the electrocardiogram as a screening tool in hypertrophic cardiomyopathy. *Congenit Heart Dis*, 2013 Sep-Oct;8(5):406-10. doi: 10.1111/chd.12083. [PubMed - https://pubmed.ncbi.nlm.nih.gov/23663480/]
6. Shreffler J, Huecker M. Diagnostic Testing Accuracy: Sensitivity, Specificity, Predictive Values and Likelihood Ratios. *National Center for Biotechnology Information*. 2023 Mar 6. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing. [PubMed - https://pubmed.ncbi.nlm.nih.gov/32491423/]
7. Principles of Epidemiology in Public Health Practice. Third Edition. Atlanta : Centers for Disease Control and Prevention (CDC), 2012. pp. 3-16. Vol. 3.

Georgi Gachev,
Address: Bulgarian Academy of Sciences, Institute of Mathematics and Informatics, 8 Acad. Georgi Bonchev Str.,1113 Sofia, Bulgaria
E-mail: gachev@math.bas.bg
Georgi Gachev https://orcid.org/0000-0002-7297-0888