KHARKIV NATIONAL MEDICAL UNIVERSITY



SCREENING.

METHODS FOR SCREENING TESTS EFFECTIVENESS ASSESSMENT

Methodical developments

for teachers to conduct a practical lesson

on the course ***“Social medicine, public health (biostatistics)”***

for students in the specialty:

– 222 “Medicine”

– 228 “Pediatrics”,

– 221 “Dentistry”.

Kharkiv

2019

MINISTRY OF PUBLIC HEALTH OF UKRAINE

KHARKIV NATIONAL MEDICAL UNIVERSITY

DEPARTMENT OF PUBLIC HEALTH AND HEALTHCARE MANAGEMENT

SCREENING.

METHODS FOR SCREENING TESTS EFFECTIVENESS ASSESSMENT

Methodical developments

for teachers to conduct a practical lesson

on the course ***“Social medicine, public health (biostatistics)”***

for students in the specialty:

– 222 “Medicine”

– 228 “Pediatrics”,

– 221 “Dentistry”.

*Затверджено вченою радою Харківського національного медичного університету.*

*Протокол № 5 від 23.05.2019*

Kharkiv

KhNMU

2019

УДК 614:616–07:001.891.5–027.236(083.13)

Screening. Methods for screening tests effectiveness assessment: methodical developments for teachers to conduct a practical lesson on the course “Social medicine, public health (biostatistics)” for students in the specialty 222 “Medicine”, 228 “Pediatrics”, 221 “Dentistry” / V.A. Ognev, I.A. Chukhno, G.V. Lisova, I.S. Bielievtsova. – Kharkiv : KhNMU, 2019. – 28 p.

|  |  |
| --- | --- |
| Drafter: | Ognev V.A. |
|  | Chukhno I.A. |
|  | Lisova G.V. |
|  | Bielievtsova I.S. |

**GUIDELINES FOR TOPIC TRAINING**

**The aim of the class**: to acquaint students with screening technologies in the health care system, to study methods for assessment the effectiveness and quality of a screening test.

**Need to know:**

* ***program questions:***
* screening;
* assessment of screening results;
* requirements for screening tests;
* sensitivity and specificity of a screening test, calculation and evaluation method;
* relationship between sensitivity and specificity;
* concept of ROC analysis.

**Need to be able to:**

* put screening technology into practice of a doctor;
* evaluate screening test effectiveness;
* make a medical decision upon the results of screening.

**The form of the lesson**: a practical lesson.

**Location**: classroom of the department.

**Methodical support of the lesson:**

* methodological recommendations for studying the topic;
* methodical literature: a workbook for students (basic training);
* presentation materials;
* test tasks.

**The algorithm of the class:** After verifying presence of students teacher announces topic and goal of class, explains relevance of it’s study and ability to use in practice.

Using educational materials provided by teacher students learn material on their own topics and perform specified tasks. At the end of class teacher will ask student what issues have arisen in preparation of topic and helps to ensure a sufficient understanding of them. Then teacher moves to control students' knowledge on basic theory material.

*Applicable forms of control:* oral questioning of students, theoretical or problematic discussion, prompt controls on options-time to complete 5–7 minutes, daily control of theoretical knowledge – individual tasks or task for several options, including 3–4 theoretical questions (time to complete not more than 20 min.), execution of tests followed by analysis of answers.

After control of theoretical knowledge teacher makes general conclusion on studied subject, sums up control of theoretical knowledge of students, as well as announcing to students their received assessment and homework.

**Plan of the lesson timing duration of occupation percentage:**

|  |  |  |
| --- | --- | --- |
| 1 | Introduction into class | up to 5% |
| 2 | Studying of key themes and issues control of theoretical knowledge | 90% |
| 3 | Summarizing and announcement homework theoretical knowledge | up to 5% |
|  | **Total** | 100% |

**Recommended literature**

**Basis literature**

1. Біостатистика / за заг. ред. чл.-кор. АМН України, проф. В.Ф. Москаленка. – К. : Книга плюс, 2009. − С. 12-31.

2. Социальная медицина и организация здравоохранения / под общ. ред. Ю.В. Вороненка, В.Ф. Москаленко. – Тернополь : Укрмедкнига. 2000. –   
С. 23-32.

3. Социальная гигиена и организация здравоохранения / под ред. Н.Ф. Серенко, В.В. Ермакова. – М. : Медицина, 1984. – С. 102-104.

4. Тестовые задачи по социальной медицине, организации здравоохранения и биостатистике : учеб.пособ. для студентов мед. ф-тов / под ред. В.А. Огнева. – Харьков : Майдан, 2005. – С. 9-14.

5.Лекционный курс кафедры.

**Additional literature**

1. Альбом А. Введение в современную эпидемиологию / А. Альбом, С. Норелл. – Таллинн, 1996. – 122 с.

2. Власов В.В. Введение в доказательную медицину / В.В. Власов. – М. : Медиа Сфера, 2001. – 392 с.

3. Герасимов А. Н. Медицинская статистика / А.Н. Герасимов. – М. : ООО «Мед.информ. агентство», 2007. – 480 с.

4. Зайцев В.М. Прикладная медицинская статистика / В.М. Зайцев, В.Г. Лифляндский, В.И. Маринкин. – СПб. : ООО «Изд-во ФОЛИАНТ», 2003. – 432 с.

5. Общая теория статистики: учебник / под ред. чл.-корр. РАН И.И. Елисеевой. − 4-е изд., перераб. и доп. − М. : Финансы и Статистика, 2000. − 480 с.

6. Основы доказательноймедицины / под ред. М. П. Скакун. – Тернополь : Укрмедкнига, 2005. – 244 с.

7. Реброва О.Ю. Статистический анализ медицинских данных. Применение пакета прикладных программ STATISTICA / О.Ю. Реброва.–М. : Медиа Сфера, 2002. – 312с.

8. Сергиенко В.И. Математическая статистика в клинических исследованиях / В.И. Сергиенко, И.Б. Бондарева. – М. : ГЭОТАР-МЕД, 2001. – 256 с.

**Information resources**

1.Население Украины. Демографический ежегодник. – К. : Госкомстат Украины –[www.ukrstat.gov.ua](http://www.ukrstat.gov.ua)

2.U.S. National Library of Medicine –Национальная медицинская библиотека США– <http://www.nlm.nih.gov/>

3.Государственная научно-педагогическая библиотека Украины им. В.О. Сухомлинского–<http://www.dnpb.gov.ua/>

4.Научная библиотека Харьковского національного медицинского университета – <http://libr.knmu.edu.ua/index.php/biblioteki>

5.Научная педагогическая библиотека им. К.Д. Ушинского Российской академии образования – <http://www.gnpbu.ru/>

6.Национальная библиотека Украины им. В.И. Вернадского –<http://www.nbuv.gov.ua/>

7.Национальная научная медицинская библиотека Украины –<http://www.library.gov.ua/>

8.Харковская государственная научная библиотека им. В.Г. Короленка – http://korolenko.kharkov.com

9.Центральная библиотека Пущинского научного центра РАН –<http://cbp.iteb.psn.ru/library/default.html>

10.Центральная научная медицинская библиотека Первого Московского государственного медицинского университетаим. И.М. Сеченова–<http://elibrary.ru/defaultx.asp>

**BASIC THEORETICAL MATERIAL**

**FOR PREPARATION FOR THE LESSON**

**1. Definition of screening test and their classification**

To ensure timely and effective intervention, modern health care system needs to receive medical information in the early stages of disease development. Currently it is possible to implement using screening technologies.

Screening tests allow distinguishing in apparently healthy population the ones who probably have the disease, and the ones who probably do not have it. Screening test is not intended to be diagnostic. Individuals with positive or suspicious results should be referred to their doctors for diagnosis and appropriate treatment. The screening initiative usually comes from the researcher, the person or organization that provides medical care, and not from the patient with the complaints. Typically, screening is aimed at chronic diseases and identifying a disease for which medical care is not yet provided. Screening allows identifying risk factors, genetic predispositions and precursors, or early manifestations of the disease.

There are different types of medical screening, each of them has its own focus.

The term screening comes from the English word “***screening***” and means “selection”, “sorting”, “separation”, it appears in its various forms and there can be:

– **screening in health care (Preventive Screening –** *Prescriptive S.*) – mass examination of individuals who do not consider themselves sick to identify hidden diseases or other conditions (risk factors for future diseases) **or screening** is a system of initial examination of groups of clinically asymptomatic individuals to identify cases of disease. A mammogram for breast cancer might serve as an example;

– **genetic screening** – using molecular biology techniques to identify mutations that are present in humans and increase the risk of the disease developing, for example, BRCA1 and BRCA2 genes, which significantly increase the risk of breast and ovarian cancer developing in women;

– **prenatal screening** is a complex of researches helped in identifying the risk of fetal malformation during pregnancy;

– **screening in microbiology** is a method used to isolate selectively target microbial species among a large microbial community;

Screening can be used in other fields, for example:

– **screening in economy** is checking the creditworthiness of potential partners, their decency;

– **screening for staff provision** is one of the recruitment technologies;

– **virtual screening** is computational procedure, which includes an automated review of chemical compounds database and selection of those for which the presence of the desired properties is predicted.

In addition, screening can be:

– **mass screening** (*Mass S.*) which means screening of the entire population;

– **multiple or multiphasic screening** (*Multiple or multiphasic S.*) involves the use of different screening tests simultaneously;

– **systematic (non-selective) screening** is performed for all individuals in a certain population, for example, ultrasound screening of chromosomal pathology, which is carried out in the first trimester of pregnancy. The population for this screening is all pregnant women, without exception;

– **selective screening** is performed among persons exposed to certain risk factors that can cause a disease. An example of such a screening is the examination of health workers for hepatitis B and C, HIV, syphilis, because representatives of these professions are in contact with the biological fluids of potentially sick people and, accordingly, have an increased risk of infection with these infectious diseases;

– **selective screening** is performed in the absence of symptoms, but with one or more risk factors for the development of the expected disease, for example, indications of the next of kin diseases, characteristics of lifestyle or belonging to the population with a high prevalence of the corresponding disease, etc.

We will mostly consider screening in the health care system when studying the subject of social medicine, public health.

**2. Purpose, objectives**

**and requirements for screening technologies**

The most frequent **purpose of screening** in medicine is detection and coverage of medical control for each case requiring medical intervention.

Connecting to this the most important practical task of screening in medicine is the identification of diseases at an early stage. This problem can be solved only if there are 2 conditions: the disease should have a preclinical period of appropriate duration and the presence of an appropriate diagnostic test also should be.

In case of a very short preclinical period during a mass examination, the probability of detecting a case of disease at an early stage will be small.

One should be remembered that there are no ideal screening tests yet. That’s why national and international organizations develop requirements for screening tests, in order to reduce diagnostic errors.

Thus, the UK National Screening Committee believes that screening test should be:

1. simple to perform, safe for the patient’s health, accurate and reliable;
2. normal distribution of the values obtained as a result of the test in the examined population should be known. Also an acceptable threshold level of the test values at which screening result will be considered positive should be established. The threshold **screening level** is the limit of a “**norm**” or the **cutoff** point beyond which the screening test is considered positive *(author’s note: The threshold screening level is also called cutoff value level (cut off point for sick or healthy person), etc.)*;
3. a test should be acceptable for the examined population;
4. a screening of gene diseases should be performed only on those diseases for which all possible gene mutations causing this disease can be diagnosed. If it is impossible to diagnose all gene mutations, screening of this gene disease should not be performed.

**WHO Expert Committee** on Health Statistics in the XI report (Geneva, 1968) offered its requirements. Therefore, a screening test according to WHO test should be:

1. reliable, i.e. provide a measurement of what is to be measured;
2. sufficiently accurate, the accuracy corresponds to the proportion of correct test results in the total number of results – both positive and negative. The required degree of accuracy depends on the studied objectives;
3. convenient, simple, cheap, affordable, well-reproducible by the examined individuals;
4. accuracy and repeatability characterize the effectiveness of a screening test.

**3. Concepts of screening**

**test and diagnostic examination**

The concept of screening test is completely different from the concept of diagnostic examination. According to the results of screening test, it is impossible to make an accurate diagnosis and assign a treatment. Screening test will allow dividing the examined individuals into 2 groups. The first group will consist of probably sick individuals, and the second group will consist of individuals for whom the disease is likely to be ruled out. Diagnostic study allows the doctor to conduct clinical interventions.

The main features and differences between the survey mass examination (screening) and diagnostic (clinical) examination are reflected in Table 1.

Screening test result can be either positive (probability of the disease presence), or negative (norm), and the true disease can be present or absent both with a positive and negative test result. Thus, there are 4 possible interpretations of the test results – 2 true and 2 false (see Table 2). The tables that represent these results are called ***contingency tables*** or 2x2 tables or four-field tables. Contingency table is a means of joint distribution of variables representation, designed to study the relationship between them. The contingency table is the most universal means for studying statistical relationships, since it can only represent absolute numbers with any level of measurement.

Regardless of where screening is applied, it has a number of key characteristics that assess screening test. These include efficiency or accuracy factors (sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), repeatability (compliance indicator and repeatability indicator)), and also screening test quality (ROC-analysis) and others.

Table 1

Differences between survey mass examination (screening)

and diagnostic (clinical) examination

|  |  |  |
| --- | --- | --- |
|  | Survey mass examination | Diagnostic clinical examination |
| 1. | Applied to population examinations | Applied to people who seek medical advice |
| 2. | Cheap and simple | Expensive and sometimes very difficult |
| 3. | Less reliable | Provides a reasonable diagnosis |
| 4. | Is not a basis for treatment prescription | Based on examination results, treatment may be initiated |
| 5. | Performed without medical indications | Performed according to medical indications |

Table 2

The ratio between the diagnostic test results

and the presence of a disease

|  |  |  |
| --- | --- | --- |
| Test Result | True Disease Status | |
| Diseased | Nondisased |
| Positive | **а**  True-positive | **b**  False- positive |
| Negative | **c**  False-negative | **d**  True-negative |

**4. Indicators characterizing**

**screening test accuracy (efficiency)**

**Test sensitivity** is the ability of the test to give a reliable assessment of ***the specific disease presence*** in the examined individual.

Tests with high sensitivity should rarely mistake in assessing the health of individuals who actually have any pathology (*disease*). In other words, sensitive tests ***should not miss sick individuals***, although at the same time, almost inevitably, in some cases, the disease can be mistakenly “attributed” to healthy individuals.

Test sensitivity is measured by specific weight (*share*) of correct (*reliable*) diagnostic assessments among individuals who are known to have this disease. The share of correct diagnoses is called ***true-positive results***. The remaining unrecognized part of the obviously sick individuals is called ***false-negative results***.

***Example***. In case a screening test in a population of 100 people suffering from this or other pathology reveals 95% of patients, and in 5% it gives a false result, then its sensitivity will be equal to 95%, respectively.

Test sensitivity is determined by the following formula (1):

(1)

where, **Se** is test sensitivity;

**a** is true-positive result;

**c** is false-negative result.

**Test specificity** is its ability to give a reliable assessment of the specific disease ***absence*** in the examined individual. Usually in this case, they say, an individual is healthy, meaning by this the absence of a certain disease.

Highly specific tests, as a rule, do not assign sick individuals to healthy ones, but may miss some cases with mild, atypical course of the disease. Test specificity is measured by the share of reliable assessments of the disease absence among obviously healthy individuals. This share of assessments is called ***true-negative results***, and the share of healthy individuals, which were mistakenly regarded as sick ones, are referred to as ***false-positive results***.

***Example***. In case a screening test in a population of 100 completely healthy individuals identified 94% as healthy, and 6 mistakenly assigned to the group with the disease, then its specificity will be equal to 94%, respectively.

Test specificity is determined by the following formula:

(2)

where, **Spe** is test specificity;

**b** is true-negative result;

**d** is false-positive result.

**The main purpose of diagnostic test** is to make a diagnosis, so it is necessary to know the probability to what extent the test allows to make the correct error-free diagnosis. *Sensitivity and specificity* do not give us this information. Instead of them, we need to analyze the so-called ***predictive values***: Test predictive value may be for both positive and negative results.

***Positive predictive value***, PPV is the probability of having a disease with a positive (pathological) test result or the share of patients with positive test results that were correctly diagnosed.

The following formula is used for determining positive predictive value (3):

(3)

where, **PPV** is positive predictive value;

**a** is true-positive result;

**b** is false-positive result.

***Negative predictive value***, NPV is the probability of a disease absence with a negative (normal) test result or the share of patients with negative test results that were correctly diagnosed.

The following formula is used for determining negative predictive value (4):

(4)

where, **NPV** is positive predictive value;

**c** is true-negative result;

**d** is false-negative result.

The factors determining screening test predictive value include the following:

1. screening test sensitivity;
2. screening test specificity;
3. prevalence of detectable disease in the studied population.

**Prevalence** is defined as the ratio of the number of individuals with the disease presence (or any other condition) to the entire population studied.

Prevalence is called ***a priori*** (pretest) probability, i.e. this is the probability to detect a disease before the test results are known. The predictive value is called ***a posteriori*** (post-test) probability of the disease.

The formula linking together sensitivity, specificity and prevalence of the disease with predictive value is derived from Bayes theorem (According to R. Fletcher “Clinical epidemiology”) (5–6):

*- for a positive result*

**PPV** = (Se\*P)/[(Se\*P)+(1-Spe)\*(1-P)] (5)

*- for a negative result*

**NPV**= (1-P)\*Spe / [(1-P)\*Spe+(1-Se)\*P)] (6)

denote, **PPV** is positive predictive value;

**NPV** is positive predictive value;

**Se** is sensitivity;

**Spe** is specificity;

**P** is prevalence.

The more ***sensitive*** is the test, the higher is its ***negative*** predictive value (i.e., increases the probability that the negative test results reject the disease presence).

On the contrary, the more ***specific*** is the test, the higher is its ***positive*** predictive value (i.e., increases the probability that positive test results confirm the proposed diagnosis).

Interpretation of positive or negative predictive value of the test result varies depending on the disease prevalence.

In case ***positive*** results of even a highly specific test are obtained in a population with a low disease prevalence, then they will turn out to be mostly ***false-positive*** ones.

In a population with no disease being studied, all positive results will be false-positive ones. Thus, when the disease prevalence tends to zero, positive predictive value also tends to zero.

In case ***negative*** results of a highly sensitive test are obtained in a population with a ***high disease prevalence***, many of them are likely to be ***false-negative*** ones. In a population where ***everyone*** has a disease, all the negative results of even a highly sensitive test will turn out to be ***false-negative*** ones. When prevalence tends to 100%, negative predictive value tends to zero. Moreover, the indicated relationships can be illustrated by analyzing the table below (Table 3). By fixing sensitivity and specificity at a constant level, and changing prevalence.

Table 3

An example of the relationship of positive predictive value (PPV)

with sensitivity, specificity and prevalence

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Test | **Cancer prevalence = 1%.** | | **Cancer prevalence = 0,1%.** | |
| cancer | no cancer | cancer | no cancer |
| Positive | 900 | 9900 | 90 | 9990 |
| Negative | 100 | 89100 | 10 | 89910 |
|  | PPV = 8,3% | | PPV = 0,9% | |

Surveyed population was 100,000 people; test sensitivity was 90%; test specificity was 90%.

If among 100,000 people, cancer is found in 1% of cases, this means that 1,000 people have cancer, and 99,000 people will not have the disease. Provided that the screening test has 90% sensitivity and 90% specificity.

A screening test that has such sensitivity and specificity will detect 900 patients out of every 1,000 of truly sick ones, but at the same time, it will assign 9,900 diseases to healthy ones. Thus, for the share of people with a positive result who are truly sick, PPV will be 900/10,800 or 8.3%.

If we use the same test on the population with the disease prevalence of 0.1%, then PPV will decrease to 0.9% – 111 false-positive results for each true case of cancer detection.

**5. Indicators characterizing**

**screening test repeatability**

***Test repeatability*** is its ability to measure ***in the same way*** any phenomena, processes, or states in a ***series*** of repeated measurements. Absolutely identical values of any health parameters after repeated examinations are relatively rare. The reasons for indicators *variability* are associated with true (*objective, biological*) and subjective variability.

***True variability*** of the results is associated with vital activity process peculiarities in the organism of the examined individual. It is known that even in healthy individuals, many indicators vary over a short period of time between studies.

***Subjective variability*** is explained by personnel errors or test errors (*equipment*).

The extent to which personnel errors can affect the results variability is clearly shown by checking the ability to measure blood pressure in special simulators that set specific and constant parameters of blood pressure. Almost always, persons who have not undergone a special training, differently measured parameters of blood pressure. Even one doctor in a series of measurements showed significant variable results. An even greater variation in the measurement results of a single indicator is observed when using different technical means, for example, different blood pressure measurement devices.

Subjective variability produces *random and systematic measurement errors*. To assess the degree of test repeatability, a series of tests is performed studying the variability of the results obtained.

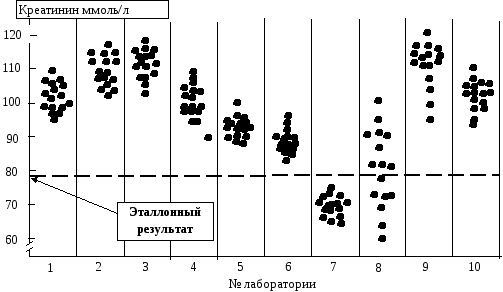
To minimize subjective variability, appropriate personnel participating in an epidemiological study should be carefully trained. First personnel training provides for training on the features of appropriate standardized methods usage. Highly qualified specialists should carry out training.

The final check and revision of the received skills occurs when applying the selected examination method to a specially selected group of individuals. In this case “*specially selected*” means that they will be similar to those who will be examined in a planned epidemiological study in terms of individual characteristics.

In addition to training in the method technique, personnel should be trained regarding standard conditions of the method application if necessary, for example, room temperature, body position of the examined individual, his/her physical activity, etc.

Subjective variability is especially great when personnel errors are combined with errors of the method (*test*).

Fig. 1 shows the results of biochemical analysis of creatinine concentration in blood. As follows from the presented data, most laboratories overestimate the results compared to the reference. In this case, laboratories No. 8 and 9 show a large variance of the results obtained.



|  |  |
| --- | --- |
| Креатинин ммоль/л | Creatinine μmol/L |
| Эталонный результат | Reference result |
| № лаборатории | Laboratory No. |

Fig. 1. Creatinine concentration in one portion of blood according to the results of 10 Swedish laboratories having analyzed the same sample 16 times. (Source: A.Albom, S. Norell. “Introduction to modern epidemiology” 1996).

The results of the study show how important it is to standardize test methods and conditions for their performance.

The results repeatability is assessed according to compliance and repeatability indicators when comparing data from 2 studies conducted under the same conditions. The formulas for their calculation are given below (7–8).

(7)

(8)

where, **a** is true-positive result;

**b** is false-positive result;

**c** is true-negative result;

**d** is false-negative result.

The German authors Bothing et. all have developed the estimation scale of repeatability indicators given in Table 4.

Table 4

Estimation scale of repeatability indicators

|  |  |  |
| --- | --- | --- |
| **Value** | **Compliance**  **indicator in %** | **Repeatability**  **indicator in %** |
| Good | 90–100 | 75–100 |
| Mediocre | 75–89 | 50–74 |
| Unsatisfactory | 74 and less | 49 and less |

Table 5 may also be used to assess the main characteristics of screening technologies.

Table 5

Summary table

of the main characteristics of screening tests

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Indicator | True-positive  (a) | False-positive  (b) | False-negative  (c) | True-negative  (d) | Calculation formula |
| Sensitivity | As much as possible |  | As less as possible |  |  |
| Specificity |  | As less as possible |  | As much as possible |  |
| Compliance and repeatability indicator | As much as possible | As less as possible | As less as possible | As much as possible |  |

**6. ROC analysis and its features**

**ROC curve** (*receiver operating characteristic*) is a diagram that allows to evaluate the quality of **binary classification**, displays the ratio between the percentage of objects in the total number of carriers of a feature, correctly classified as carrying a feature, (*true-positive rate*, TPR, which is abovementioned as classification algorithm *sensitivity*) and the percentage of objects in the total number of objects that do not carry a feature, incorrectly classified as carrying a feature (*false-positive rate*, FPR, the value 1-FPR is mentioned as classification algorithm *specificity*) by varying the threshold of the decision rule.

ROC curves were first used in signal processing theory in the USA during the World War II to improve the recognition quality of enemy objects (Japanese planes) using a radar signal.

Subsequently, ROC curves have been widely used in medical diagnostics, often in epidemiology and medical research, and are mentioned in the same context as evidence-based medicine.

Classifications analysis using ROC curves is called **ROC analysis**.

The task of classification is to assign previously unknown entities to one or another class. An example of such a task might be a diagnosis based on the results of a medical examination. In this case, there are two classes of results: positive and negative. Then at the output of the classifier, we will have four different situations:

1. If the result of the classification is *positive*, and the true value is also *positive*, then we are talking about a *true-positive value (TP)*.

2. If the result of the classification is *positive*, but the true value is *negative*, then we are talking *about a false-positive value (FP)*.

3. If the result of the classification is *negative*, and the true value is also *negative*, then we are talking about a *true-negative value (TN)*.

4. If the classification result is *negative*, but the true value is *positive*, then we are talking about a *false-negative value (FN)*.

Returning to the example of a test for any disease, let us suppose that a doctor needs to establish the diagnosis of cancer or its absence on the basis of medical research. Then:

– *true-positive, TP* – the patient has cancer, the diagnosis is positive;

– *false-positive, FP* – the patient is healthy, the diagnosis is positive;

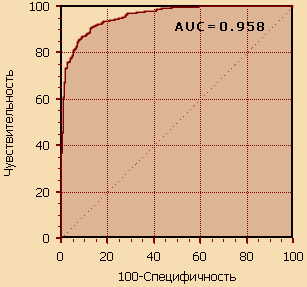
– *true-negative, TN* – the patient is healthy, the diagnosis is negative;

– *false-negative, FN* – the patient has cancer, the diagnosis is negative.

Four possible outputs can be formulated and arranged in the form of a previously considered contingency table having size 2×2.

Then the value of Sen=TP/(TP+FN), the ability of the algorithm to “see” sick individuals is **sensitivity**, and Spe=TN/(TN+FP) is **specificity**, the ability of the algorithm not to take healthy individuals as ill ones.

It happens that the classifier produces not 2 options “healthy or sick”, but a different number, for example, 5: “obviously healthy” – “most likely to be healthy” – “uncertainly” – “most likely to be sick” – “obviously sick”. This is better, but still the set of decisions made is final, and often, all the same, it is binary: should a patient be sent for an additional examination?

ROC curve is built as follows:

1. Se sensitivity and Sp specificity values are calculated for each cutoff value (cutoff point for sick or healthy or threshold **screening level**), which varies systematic from 0 to 1 with dx step (for example, 0.01). Alternatively, each subsequent value of the sample in selection may be the threshold.

2. A dependence diagram is built: along the Y-axis, Se sensitivity is plotted, along the X-axis – 100% – Sp (one hundred percent minus specificity), or, equivalently, FPR – the percentage of false positive cases.

For an ideal classifier, ROC curve passes through the upper left corner, where the percentage of true-positive cases is 100% or 1.0 (ideal sensitivity), and the percentage of false-positive examples is zero. Therefore, the closer the curve is to the upper left corner, the higher the predictive ability of the model. On the contrary, the smaller the bend of the curve and the closer it is to the diagonal line, the less effective is the model. The diagonal line corresponds to “useless” classifier, i.e. complete indistinguishability of two classes.

**7. Overall assessment of screening test effectiveness**

A visual comparison of 2 or more ROC curves does not always reveal the most effective model. In this connection, an evaluation of the area under the curves is a kind of quantitative method for ROC curves comparing. Theoretically, it varies from 0 to 1.0, but since the model is always characterized by a curve located above the positive diagonal, it is usually said about from 0.5 (“useless” classifier) to 1.0 (“ideal” model). This estimate can be obtained directly by calculating the area under the polyhedron, bounded on the right and below by the axes of coordinates and on the upper left by experimentally obtained points. The numerical value of the area under the curve is called AUC.

With large assumptions, we can assume that the greater is AUC, the better predictive power the model has. However, you should know that:

– AUC indicator is intended rather for comparative analysis of several models;

– AUC does not contain any information about model sensitivity and specificity.

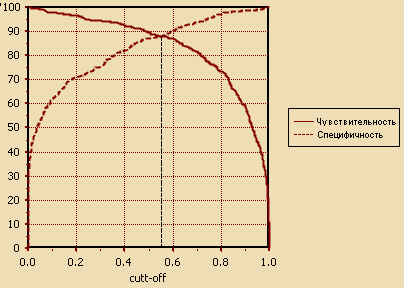
The following expert scale for AUC values is sometimes given in the literature, according to which model quality can be judged (Table 6):

Table 6

Expert scale for AUC values

|  |  |
| --- | --- |
| **AUC interval** | **Model quality** |
| 0,9-1,0 | Perfect |
| 0,8-0,9 | Very good |
| 0,7-0,8 | Good |
| 0,6-0,7 | Mediocre |
| 0,5-0,6 | Unsatisfactory |

The ideal model has 100% sensitivity and specificity. However, in practice this cannot be achieved; moreover, it is impossible to simultaneously increase both the sensitivity and the specificity of the model. The compromise is found with the help of cutoff value, since the threshold value affects Se and Sp. One can talk about the task to find the *optimal cutoff value*.

Cutoff value is necessary in order to apply the model in practice: to assign new examples to one of two classes. To determine the optimal threshold, one needs to specify a criterion for its determination, since different tasks have their own optimal strategy. The criteria for cutoff value selecting may be the following:

– the requirement for the minimum value of the model sensitivity (specificity). For example, one needs to ensure the sensitivity of the test at least 80%. In this case, the optimal threshold will be the maximum specificity (sensitivity), which is achieved at 80% (or a value close to it “on the right” because of the series discreteness) of sensitivity (specificity);

– the requirement for the maximum total sensitivity and specificity of the model;

– the requirement for a balance between sensitivity and specificity, i.e. when Se≈Sp;

– the optimal threshold can be intersection point of two curves when the cutoff value is plotted along the X axis, and the model sensitivity or specificity is plotted along the Y axis.

**8. Questions answered by diagnostic test characteristics**

1. *Sensitivity* – to which extent the test is good to identify patients having this condition?

2. *Specificity* – to which extent the test is good to exclude properly patients not having this condition?

3. *Positive predictive value of the test result* – if an individual has a positive test, what is the probability that he/she actually has the disease?

4. *Negative predictive value of the test result* – if an individual has a negative test, what is the probability that he/she does not actually have this disease?

5. *Repeatability indicator* – how close are the individual values in a series of results of repeated (parallel) measurements, or dispersion degree of results in different (repeated) studies comparing to an average value.

6. *Accuracy index* – what part of all tests has provided correct results (i.e. true-positive and true-negative results for all)?

7. *Likelihood ratio of a positive test* – how much more likely is it that the test will be positive in an individual with a disease compared to a healthy one?

**PRACTICAL TASK**

In the settlement, located in the immediate vicinity of a large industrial complex, a selective screening test was conducted among the population in order to identify angina pectoris. 337 people were examined, of which 17 people had positive results and 320 had negative results. The detailed distribution of screening test is shown in the table.

It is necessary to:

* + calculate accuracy indicators (sensitivity, specificity, positive and negative predictive value of the result) and repeatability (compliance and repeatability indicator);
  + conduct ROC analysis (assess screening test quality);
  + make conclusions.

In the course of students’ independent work, the teacher answers the questions that arise and monitors the correctness of the task performance. After the end of independent work, the teacher checks the task performance.

Table

Screening test results distribution

among the population in order to identify the results of angina pectoris

|  |  |  |  |
| --- | --- | --- | --- |
| Screening test | Medical enquiry | | Total |
| Present | Absent |
| Positive | 10 (a) | 7 (b) | 17 |
| Negative | 34 (c) | 286 (d) | 320 |
| Total | 44 (a+c) | 293 (b+d) | 337(a+b+c+d) |

Sensitivity indicator:

Specificity indicator:

Positive predictive value (PPV):

Negative predictive value (NPV):

Compliance indicator:

Repeatability indicator:

**2. ROC** analysis of screening test

|  |  |
| --- | --- |
| Чувствительность | Sensitivity |
| 100-специфичность | 100-specificity |

**Conclusions:**

1. The screening test being analyzed has a low sensitivity (22.7%), positive predictive value (58.8%) and a repeatability indicator (19.6%), at the same time a rather high specificity (97.6%), negative predictive value (58.8%) and compliance rate (87.8%).

2. For the ideal classifier, ROC curve diagram passes through the upper left corner, where the percentage of true-positive cases is 100% or 1.0 (ideal sensitivity), and the percentage of false-positive examples is zero. Therefore, the closer the curve is to the upper left corner, the higher the predictive ability of the model. On the contrary, the smaller the bend of the curve and the closer it is to the diagonal line, the less effective is the model. The diagonal line corresponds to “useless” classifier, i.e. complete indistinguishability of two classes.

In this case, due to the lack of step-by-step calculation of sensitivity and specificity, it is impossible to build a qualitative ROC curve, nor can we calculate the area under the curve called AUC, we can only assume that the curve will pass above the diagonal line, but this is not enough to talk about effective screening test.

**TEST TASKS**

|  |  |  |
| --- | --- | --- |
| 1. | To ensure effective intervention a modern health care system needs to receive accurate medical information about public health, which is currently possible with the help of screening technologies. Choose the correct definition of “screening”. | |
|  | А | All concepts are true |
|  | В | Diagnostic examination using modern expensive methods of detecting diseases among the population |
|  | \* С | Mass examination of individuals who do not consider themselves sick, to identify hidden diseases or other conditions |
|  | D | Medical examination of the population applying to a medical institution |
|  | E | One-time survey of the population or separate groups in order to identify diseases |
| 2. | The WHO Expert Committee on Health Statistics in the XI report (Geneva, 1968) offered screening test requirements. Specify their basic requirements for diagnostic tests. | |
|  | А | Accuracy |
|  | \* В | All listed above |
|  | С | Credibility |
|  | D | Efficiency |
|  | Е | Simplicity, low cost |
| 3. | Tests with high accuracy should rarely make mistakes in assessing the health of individuals; in this case, sensitivity plays an important role. What is diagnostic test sensitivity? | |
|  | А | The ability of a test to give accurate results |
|  | В | The percentage of individuals with a condition defined as negative |
|  | С | The percentage of individuals with a condition defined as negative in case the disease is absent |
|  | \* D | The probability of a positive diagnostic test in case the disease is present |
|  | Е | The share of individuals with a negative test result in the population with the studied disease |
| 4. | Test technologies should assess the state of public health as accurately as possible; in this case, specificity plays an important role. What is diagnostic test specificity? | |
|  | А | The ability of a test to give accurate results |
|  | В | The percentage of individuals with a condition defined as negative in case the disease is present |
|  | С | The percentage of individuals with a condition defined as positive |
|  | \* D | The probability of a negative diagnostic test result in case the disease is absent |
|  | Е | The proportion of individuals with a positive test result in the population with the studied disease |
| 5. | The main purpose of a diagnostic test is to make a diagnosis, so one needs to know the probability to which extent the test allows to make the correct error-free diagnosis. What does positive predictive value mean? | |
|  | А | The percentage of individuals with a negative test result who are actually healthy |
|  | В | The percentage of individuals with a positive test result who are actually healthy |
|  | С | The probability of the disease absence in case of a negative (normal) test result |
|  | D | The probability the disease presence in case of a negative (normal) test result |
|  | \* Е | The probability of the disease presence in case of a positive (pathological) test result |
| 6. | The purpose of a diagnostic test is to make a diagnosis, so one needs to know the probability of to which extent the test effectively recognizes the prognosis. What does negative predictive value mean? | |
|  | А | All answers are correct |
|  | В | The percentage of individuals with a positive test who are actually sick |
|  | С | The probability of a negative diagnostic test in case the disease is present |
|  | \* D | The probability of the disease absence in case of a negative (normal) test result |
|  | Е | The probability of the disease presence in case of a positive (pathological) test result |
| 7. | One of the requirements of the WHO Sanitary Statistics Committee experts for screening tests is test repeatability. What does repeatability of a diagnostic test mean? | |
|  | А | The probability of the disease presence, subject to a known test result  The ability of the test not to miss patients who have a disease |
|  | В | The ability of the test not to miss patients who have no diseases |
|  | С | The ability of the test to give objective, reliable results |
|  | D | The probability of the disease presence, subject to a known test result |
|  | \* Е | The probability that the repeated measurements of a stable phenomenon will give the same result |
| 8. | Screening tests allow distinguishing in apparently healthy population the ones who probably have the disease, and the ones who probably do not have it. Individuals with positive or suspicious results should be referred to their doctors for diagnosis. List the indicators by which the accuracy of the diagnostic test is measured. | |
|  | А | Absence of I and II order errors |
|  | В | Specificity, repeatability, absence of random errors |
|  | С | Sensitivity, repeatability, lack of systematic errors |
|  | \* D | Sensitivity, specificity, test predictive value |
|  | Е | Test predictive value, absence of random and systematic errors |
| 9. | To study the prevalence of arterial hypertension in the population, a one-stage epidemiological study was conducted using a screening test with 90% sensitivity and 60% specificity parameters. What share are actually healthy among those who were classified as healthy according to the test results? | |
|  | А | 10 |
|  | В | 40 |
|  | \*С | 60 |
|  | D | 90 |
|  | Е | 100 |
| 10. | A test used in cohort studies has 70% sensitivity, 60% specificity, a good repeatability and 75% compliance rate. What will the share of actually sick among those surveyed who were classified as sick, according to the test results? | |
|  | А | 10 |
|  | В | 60 |
|  | \*С | 70 |
|  | D | 75 |
|  | Е | 80 |
| 11. | In the practice of a doctor when comparing several tests for screening a particular disease, there often arises the question which test is the most effective. ROC curves are used for this purpose. What is a kind of quantitative method for comparing ROC curves? | |
|  | А | Evaluation of the area above the curves |
|  | \* В | Evaluation of the area under the curves |
|  | С | Likelihood ratio |
|  | D | Repeatability indicators |
|  | Е | Sensitivity and specificity |
| 12. | To analyze the quality of screening tests, in some cases it is necessary to use binary classifiers that reflect the ratio between the percentage of objects in the total number of carriers of a feature correctly classified as bearing a feature and the percentage of objects in the total number of objects that do not carry a feature, incorrectly classified as carrying a feature. Which of the following allows assessing the quality of a binary classification? | |
|  | А | Absence of random and systematic errors |
|  | В | Negative predictive value |
|  | С | Positive predictive value |
|  | \*D | ROC curve |
|  | Е | Specificity and sensitivity |

**CONTROL QUESTIONS**

1. What is a screening test? The main goal and objectives of screening technology.

2. What are the types of screening test, give examples?

3. List the main requirements of experts to screening tests.

4. What are the differences between the concept of screening test and diagnostic examination?

5. What are contingency tables, and their interpretation?

6. What does sensitivity and specificity of screening test mean?

7. How do you understand positive and negative predictive value in screening technologies, their calculation?

8. Indicate the presence of a relationship between pathology sensitivity, specificity and prevalence.

9. What is compliance indicator and repeatability indicator, their calculation method.

10. ROC analysis and its task.

11. The order of ROC curves construction and their assessment.

**CONTENT**

|  |  |
| --- | --- |
| Guidelines for topic training ……………………………………………….. | 3 |
| Basic theoretical material for preparation for the lesson…………………… | 6 |
| 1. Definition of screening test and their classification…………..... | 6 |
| 2. Purpose, objectives and requirements for screening technologies………………………………………………………………… | 7 |
| 3. Concepts of screening test and diagnostic examination……..…  4. Indicators characterizing screening test accuracy (efficiency)....  5. Indicators characterizing screening test repeatability……..……  6. ROC analysis and its features……………………………..……  7. Overall assessment of screening test effectiveness…………..…  8. Questions answered by diagnostic test characteristics……….. | 8  9  13  15  17  18 |
| Practical tasks ……………………………………………………………… | 20 |
| Test tasks…………………………………………………………………… | 23 |
| Control questions…………………………………………..……………….. | 27 |

*Educational publication*

**SCREENING.**

**METHODS FOR SCREENING TESTS**

**EFFECTIVENESS ASSESSMENT**

Methodical developments for teachers   
to conduct a practical lesson on the course:

***“Social medicine, public health (biostatistics)”***

for students in the specialty:

222 “Medicine”, 228 “Pediatrics, 221 “Dentistry”.

|  |  |
| --- | --- |
| Drafters: | ***Ognev Viktor*** |
|  | ***Chukhno Inna*** |
|  | ***Lisova Ganna*** |
|  | ***Bielievtsova Iana*** |

*Translated by Chukhno I., Maazouzi Hind, Ivahnenko D.*

Responsible to sign-off *Ognev V.*

Формат А5. Ризографія. Ум.друк. арк. 1,75.

Тираж 100 прим. Зам. № 19-33765.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

Редакційно-видавничий відділ

ХНМУ, пр. Науки , 4, м. Харків, 61022

izdatknmu@mail.ru, izdat@knmu.kharkov.ua

Свідоцтво про внесення суб’єкта видавничої справи до Державного реєстру видавництв,

виготівників і розповсюджувачів видавничої продукції

серії ДК № 3242 від 18.07.2008 р.