



COVID-19: About the Tests' Accuracy, Sensitivity and Specificity Again

Peter DINEV, Iliya TSENEV, Ivaylo IVANOV, Krum KATSAROV

Kachestvo 21 Vek Foundation, Sofia, Bulgaria
e-mail: office@kachestvo-21.com

Abstract

What should we know about testing, the basic test characteristics, the performance of the COVID-19 tests used? How to set the target before testing and how to read the results? These questions are all too important given that: the number of COVID-19 cases confirmed by testing still remains significantly less than the number of reported infections; the number of COVID-19 tests that can be done also remains limited by the supply of consumables such as tampons and reagents, as well as the shortage of trained people and protective equipment.

Keywords: COVID-19, data analysis

COVID-19: Отново за точността, чувствителността и специфичността на тестовете

Петър ДИНЕВ, Илия ЦЕНЕВ, Ивайло ИВАНОВ, Крум КАЦАРОВ

„И не можем да спрем тази пандемия, ако не знаем кой е заразен. Имаме просто съобщение за всички страни: тествайте, тествайте, тествайте! Тествайте всеки съмнителен случай!

„За да победим, трябва да атакуваме вируса с агресивни и целенасочени стратегии – да тестваме всеки подозрителен случай, да изолираме всеки потвърден, да откриваме и поставяме под карантина всички в близък контакт.“

Tedros Adhanom Ghebreyesus, СЗО

„Силната интуиция е много по-мощна от слабия тест!“
Siddhartha Mukherjee

Въведение.

Повод за настоящата статия стана търсенето на отговор на поставения в заглавието въпрос, който е пряко свързан с обработването на данни от тестовете за COVID-19, най-често чрез бинарна класификация с помощта на регресионен анализ и създаването на статистически модели. Този тип класификация разделя случайната извадка на две подмножества: да/не, вярно/невярно, правилно/неправилно. Това е най-простият вид класификация, използвана при обработването и оценяването на резултатите от тестване (изпитване, изследване), която съвсем не е чужда на инженерите от областта на машиностроенето, електротехниката, автоматизацията и електрониката. Съгласни сме, че не сме здравните специалисти или епидемиолозите, чиито становища трябва да се тълкуват като професионални съвети. Но в едно сме сигурни, че читателите ще ни разберат и ще могат да си съставят свое независимо мнение по един толкова актуален световен проблем.

За коронавирусната болест

Коронавирусната болест (*Corona Virus Disease 2019*; COVID-19) е инфекциозно заболяване, което се причинява от вируса SARS-CoV-2 (*Severe Acute Respiratory Syndrome CoronaVirus 2*). Заболяването е установено за първи път през декември 2019 година в Ухан, столицата на провинция Хубей (Китай), и оттогава се разпространява в световен мащаб. На 11 март 2020 год., Световната здравна организация (СЗО) обявява началото на продължаващата пандемия от *Covid-19*. Към 01 май 2020 год., 3,27 милиона случая на заболяване са регистрирани в 187 държави и територии, в резултат на което са починали над 233 000 души. Повече от 1,02 мил. души вече са се възстановили от болестта.

За тестването, вероятността преди теста и неверните положителни и отрицателни резултати

Неверните положителни и отрицателни резултати

Тестването на COVID-19 може да идентифицира вируса на SARS-CoV-2 основно чрез два вида методи, които: *първо*, откриват пряко самото присъствие на вируса (PCR тестове в реално време) и *второ*, откриват антителата, произведени в отговор на инфекцията. Откриването на различни антитела може да се използва, както за диагностика, така и за наблюдение върху популацията. Тестовите с антитела показват колко хора са имали заболяването, включително тези, които са били безсимптомни или симптомите им са били незначими. По резултатите от този тест би могло да се проследи заболяването сред популацията, да се открият огнища на заразяване, да се определи смъртността и нивото на „имунитета на стадото“. Въпреки това, продължителността и ефективността на този имунен отговор все още не са ясни и честотата на „*неверните положителни*“ и „*неверните отрицателни*“ резултати трябва да бъдат основателно включени в интерпретацията на резултатите от теста.

Към края на април страните, които са публикували свои данни от проведени тестови изследвания, са обхванали едва 1,4 % от населението си. Няма държава, която да е тествала повече от 14 % от своето население. Съществуват различия в това, колко тестове са направени в различните страни, Табл. 1.

Таблица 1. Статистически данни за осъществени тестове от едни от най-засегнатите държави от COVID-19. За сравнение са дадени данни за България.

Страна	Данни към дата	Тествани		
		бр.	Позитивни бр.	%
САЩ	01 май	6 553 449	1 095 681	16,70
Русия	28 април	3 723 807	114 431	3,10
Китай (Гуангдонг)	14 април	3 650 000	1 954	0,05
Китай (Ухан, Хубей)	23 април	1 615 987	51 850	3,20
Китай (Хонг Конг)	27 април	154 989	1 037	0,67
Германия	28 април	2 547 052	157 641	6,20
Италия	01 май	2 053 425	207 428	10,10
Испания	27 април	1 345 560	229 422	18,20
ОАЕ	29 април	1 122 000	11 380	1,010
Турция	01 май	1 075 048	122 392	11,40
Великобритания	01 май	1 023 824	177 454	17,30

Индия	02 май	976 363	37 336	3,80
Южна Корея	28 април	605 890	10 761	1,80
Франция	21 април	595,154	117 324	19,70
България	29 април	45 208	1 447	3,20

От началото на пандемията лекари започват да съобщават за случаи на хора, които са били действително болни от COVID-19, но са били тествани отрицателно чрез т.нар. „*неверен* (фалшив, лъжлив) *отрицателен резултат*“, т.е. били са обявени за здрави. Не се знае със сигурност колко често тези *неверни отрицателни резултати* се срещат в хода на тестването (Табл. 1), но данни от Китай предполагат, че до 30 от всеки 100 души с COVID-19 заболяване, могат да бъдат тествани отрицателно чрез PCR тестване. Съобщава се още за случаи на хора, които действително не са били болни от Covid-19, но са били тествани положително, чрез „*неверен положителен резултат*“, и са били обявени за болни с всички последици за тях, за семействата, за работодателите и обществото.

Каква е причината? На първо място недоказаните тестове и неточните резултати. Лабораториите за обществено здраве вече се притесняват, че „лошите данни“ могат да застрашат сериозно възстановяването на САЩ от „коронавирусната“ криза. *Colin West*, лекар и професор по вътрешна медицина в клиниката *Мауо*, който е валидирал тестове за COVID-19, заключава: „*Трябва да спрем да се преструваме, че тестовете са перфектни!*“. И препоръчва основателно: този, който се тества, не трябва „*изведнъж да спре да мие ръцете си, да спре да носи маска или да прекрати физическото дистанциране*“. Според *West* проблемът е в това, че няма адекватни доказателства, че тестовете са достатъчно добри, което да облекчи политиките за социално дистанциране. Да, казва той, ще стигнем и до това, но сега е твърде рано.

Без „паспорти за имунитет“, предупреждава СЗО

Кръвните тестове, наричани още серологични тестове, са предназначени да открият дали човек е разработил антитела за борба с вируса SARS-CoV-2, причиняващ COVID-19. Обикновено антителата не се откриват до две седмици след заразяване на човек. Все още, изследователи теоретизират, дали антителата предпазват от второ заболяване, може ли да се гарантира една продължаваща във времето защита и дали е напълно ефективна, доколкото други вирусни заболявания, включително от други по-леки коронавирусни заболявания, влияят кръстосано върху резултатите. Несъмнено, ще отнеме време, за да се събере достоверна информация от нови изследвания, за да се отговори на тези въпроси. Ето защо, СЗО предупреди да не се издават т.нар. „паспорти за имунитет“ на хора, които вече смятат, че са в безопасност, въз основа на положителни тестове за антитела. В свое изявление, СЗО заяви, че няма доказателства за това, че тези, които се възстановяват от заболяването COVID-19 и развиват антитела, ще бъдат имунизирани от последваща инфекция.

Много тестове, масово тестване и проблеми

Лабораториите в САЩ вече тестват до 100 000 души всеки ден: Това все още не е достатъчно. Но служители от *Асоциацията на лабораториите за обществено здраве (APHL)* предупреждават, че тестовете за антитела на пазара са недоказани и могат да доведат до „неточни резултати“. Групата призова за по-строг надзор от страна на *Администрацията по храните и лекарствата (FDA)*, която разрешава на производителите на тестове за антитела да продават своите продукти без обичайната стъпка на подаване на данни за валидиране на теста. Повече от 100 производители на тестове са уведомили FDA за своите планове да пуснат на пазара тестове за антитела.

Компаниите или лабораториите не могат да твърдят, че FDA е разрешила тестовете, и трябва да носят отговорност при невярно заключение за това дали човек е преболедувал болестта. Повече от 250 вида тестове, обаче, вече са на пазара в САЩ. Политиката на FDA, която позволява на тези тестове за антитела да бъдат на пазара, преди тя да е ги е сертифицирала, има за цел единствено да балансира рисковете и ползите, за да бъдат посрещнати „спешните нужди на общественото здраве“. Администрацията FDA е прегледала и разрешила единствено тестовете за антитела, произвеждани от осем компании/лаборатории.

По този повод, Colin West казва, че е важно да се признаят и да се отчетат недостатъците на тестовете за антитела. Използването им за информирано вземане на решения е „наистина само толкова ефективно, колкото тестът е точен – трябва да вградим техните несъвършенства в нашите модели и в нашите публични политики“. „Очевидно е, че хората искат да знаят дали са били заразени и търсят тези тестове. Всъщност, ние не знаем колко точно е „точен“ по отношение на „неверните положителни“ и „неверните отрицателни“ резултати“ един тест, да не говорим, че не знаем дали наистина се създава имунитет след заболяването“, каза доктор Том Инглесби, директор на Центъра за здравна сигурност (Училище за обществено здраве „Johns Hopkins Bloomberg“). И продължава: „Първоначалните проблеми с тестовете за антитела са поправими, но ще отнеме време, за да се появят правилните „контроли“, и да се правят научно обосновани твърдения“.

Силната интуиция е много по-мощна от слабия тест

Известно е това, че лекарите използват своя опит, за да разпознаят симптомите, рисковите фактори и признаците, за да преценят каква е т.нар. вероятност от заболяване преди самото тестване или по-просто „вероятност преди тестване“. Значението на резултатите от теста, за конкретния човек, зависи не само от „точността“ на теста. Прогнозирането на риска от заболяване, преди самото тестване, е много важно да бъде извършено от лекаря. Това положение е описано математически от *Thomas Bayes* (1701÷1761) и по-късно формулирано от *Siddhartha Mukherjee* (1970÷) като закон: „Силната интуиция е много по-мощна от слабия тест!“.

Най-честите симптоми на COVID-19 са високата температура, умората и сухата кашлица. Някои пациенти могат да имат загуба на миризма в продължение на дни, главоболие, назална конгестия, хрема, болки в гърлото или диария. Симптомите, обикновено, са леки и започват постепенно. Някои хора се заразяват, не развиват никакви симптоми, но не се чувстват добре. Повечето хора (около 80 %) се възстановяват от болестта без да се нуждаят от специално лечение. Около 1 на всеки 6 души, се разболява сериозно от COVID-19, и развива затруднено дишане. Възрастните хора и тези с медицински проблеми, като високо кръвно налягане, сърдечни проблеми или диабет рискуват да развият сериозно заболяване. Хората с треска, кашлица и затруднено дишане задължително трябва да потърсят лекарска помощ.

Така, нека се приеме, че 100 души са изложени на риск от заразяване и имат най-често срещаните симптоми на COVID-19 – високата температура, умората и сухата кашлица. Нека, 80 от тях имат действително заболяването COVID-19. Положителният тест ще означава, със сигурност, че някой има това заболяване. Но ако тестът пропусне 30 % от тези 80 души с COVID-19, това вече ще означава, че приблизително 24 от тестваните 100 души имат „неверен отрицателен“ резултат от теста. Това е „страхотна новина“ за тези хора, та те не са болни? Едва ли в действителност е така, тези хора са болни, но се връщат на работа, в семейството, на улицата, и несъзнателно разпространяват заболяването. Това вече е много тревожно!

Тази ситуация твърде много зависи от това, *кой е тестваният*. Ако се тестват хора с по-малко симптоми, които не работят на рискови места, вероятността от заболяване или „*вероятността преди теста*“ е значително по-ниска. Нека, само десет от 100 тествани действително имат COVID-19, а 30 % от тях са пропуснати от теста, което означава, че само 3 от 100 тестваните ще имат „*неверен отрицателен резултат*“, в сравнение с 86 „*верни отрицателни*“ резултати. Така, че ако болният има по-малко симптоми и резултатът от теста е отрицателен, то лекарят може да бъде значително по-сигурен, че няма заболяване от COVID-19. *Внимание*, въпреки „*верния отрицателен*“ резултат, ако са налице типичните симптоми на заболяването, то тогава е най-безопасно да се приеме, че заболяване има!

Налични данни предполагат, че кръвните тестове за антитела имат по-малко „*фалшиво отрицателни*“ резултати от PCR- (с тампони) тестовете, но дават повече „*невярно положителни*“ резултати от тях. Това означава, че има вероятност да тествате „*положително*“ здрави хора, т.е. да ги обявите за болни. Този вид тестване е важно, за да се разбере и проследи разпространението на болестта, както и да се контролират огнищата на болестта. Те, обаче, имат ред ограничения, когато резултатите се насочат към вземане на решения за конкретни хора. *Запомнете*, дори вашият тест да е отрицателен, но имате силно изразени симптоми на COVID-19, то трябва да приемете, че заболяване има!

В същото време, според СЗО, чувствителността на подобни антигенни тестове за респираторни заболявания, като грипа, варира между 34 % и 80 %. Въз основа на тази информация, половината или повечето от заразените с COVID-19 изследвани, могат да бъдат пропуснати от такива тестове, в зависимост от групата на тестваните пациенти. Много учени дори се съмняват дали антигенният тест може да бъде достатъчно надежден във времето, за да бъде полезен срещу COVID-19.

За „невярно положителните“ и „невярно отрицателните“ резултати, и точността, чувствителността и специфичността на тестовете за Covid-19

През март 2020 година, Китай съобщава за проблеми с „*точността*“ на техните тестови комплекти. Чехия потвърждава, че 80 % от закупените тестови комплекти от Китай дават „*грешни*“ резултати. Словакия обявява закупените 1,2 мил. тестови комплекти от Китай за „*неточни*“. Турското министерство на здравеопазването „не пуска“ в употреба, закупените от Китай комплекти, тъй като имат „*висок процент на грешки*“. Обединеното кралство закупува 3,5 милиона тестови комплекта от Китай, но в началото на април обявява, че „*те са неприложими*“. На 21 април, Индийският съвет за медицински изследвания препоръчва на индийските щати да спрат да използват комплектите за бързо тестване на антитела, закупени от Китай, тъй като щатът Раджастан заявява, че тестовете, закупени от Китай, дават „*многократно по-ниска точност*“ от очакваната.

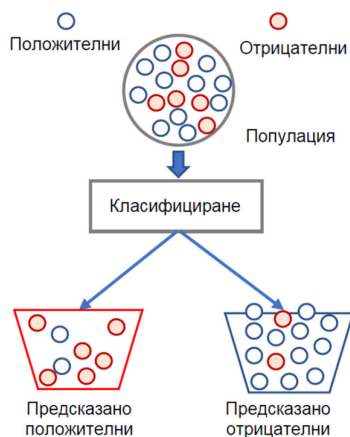
В същото време, на 29 април се появява съобщение, че кръвният тест за COVID-19 отговаря на правилата за безопасност и е **сертифициран** за използване в Европейския съюз. Тестът открива IgG-антитела, които се появяват в късните етапи на инфекцията и остават известно време, след като човекът се е възстановил. Производителят *Abbott* твърди, че тестът е 99 % точен. Веднага, обаче, *Abbott* съобщава, че тестът има 100 % **чувствителност** (95 % доверителен интервал, от 95,07 % до 100 %) за 73 положителни проби, взети 14 дни след появата на симптомите (без имунно компрометиран пациент) и 99,6 % **специфичност** (95 % доверителен интервал, от 99,05 % до 99,90 %) за 1 070 отрицателни проби.

Използваме двата случая, за да бъде поставен основателно следния въпрос: Каква е и каква трябва да бъде „точността“ на тестовете за COVID-19? Точността или кои са базовите статистически показатели на тестовете?

Базови статистически показатели на тестовете за COVID-19. Какво трябва да знаем, за да разберем проблемите с тестовете?

Тестовото изследване има за основна цел да раздели две случайни множества „скрити“ в изследваната популация – множеството на „действително положителните“ с COVID-19, или болелите, и множеството на „действително отрицателните“ без COVID-19, или здравите. Тестът, като инструмент за бинарно (да/не) класифициране, извършва тази класификация, но в общия случай двете множества се припокриват, тъй като разделя цялото множество на две подмножества – „предсказани положителни“ и „предсказани отрицателни“, но това не са действително положителните и отрицателните стойности. Появяват се още „положително отрицателни“ и „отрицателно положителни“ стойности, които остават скрити в двете класифицирани множества, фиг. 1.

Както се вижда, „класификаторът“ работи при известни „действително положителни“ и „действително отрицателни стойности“. Лекарят трябва да използва своя опит, за да разпознае *симптомите, рисковите фактори и признаците*, за да прецени „вероятността преди тестване“. Значението на резултатите от теста, отнесено за конкретен човек, зависи не само от „точността“ на предсказване на теста. Ето защо, прогнозирането на риска от заболяване, преди самото тестване, е много важно да бъде извършено от лекар.



Фигура 1.

		ДЕЙСТВИТЕЛНО СЪСТОЯНИЕ	
		ЖЕНА	МЪЖ
РЕЗУЛТАТИ ОТ ТЕСТА ЗА БРЕМЕННОСТ	Положителен резултат	Верен (действителен) положителен резултат	Неверен (фалшив) положителен резултат
		1	0
	Отрицателен резултат	Неверен (фалшив) отрицателен резултат	Верен (действителен) отрицателен резултат
		0	1

Фигура 2.

Матрица на объркване

Матрицата на объркване (*Confusion Matrix*), или смесване, представлява таблица, която се използва, за да се илюстрира приложението на избрания подход (начин, модел, технология) на класификация върху набора от тестови данни, за които са известни истинските стойности, фиг. 2.

Матрицата (таблицата) на объркване показва броя на вярно и невярно класифицираните резултати, в сравнение с действителните резултати в данните от теста, фиг. 2.

Най-простата класификация, наречена двоична класификация (*Binary Classification*), има само два класа: да/не, верен/неверен, положителен/отрицателен, правилно/неправилно, мъжки/женски, здрави/болни, и т.н. Реализирането на двоичната класификация се обобщава в матрицата на объркване, която пресича таблично прогнозираните и наблюдавани резултати в четири варианта, Табл. 2:

- Вярно (действително) положителен резултат (*True Positive*, TP): вярно предвиждане, т.е. *предвидено е „да“ и това е „да“!*
- Вярно (действително) отрицателен резултат (*True Negative*, TN): вярно предвиждане, т.е. *предвидено е „не“ и това е „не“!*
- Невярно (фалшиво) положителен резултат (*False Positive*, Type I Error, FP): невярно предвиждане, т.е. предвидено е „да“, но това е „не“!
- Невярно (фалшиво) отрицателен резултат (*False Negative*, Type II Error FN): невярно предвиждане, т.е. предвидено е „не“, но това е „да“!

Таблица 2. Матрица на объркването (грешката)

Случайна извадка от резултати, общо: Total (TP+TN+FP+FN) =			Действително състояние		Базови статистически показатели
			Болни от Covid-19	Здрави (не са болни)	
Резултати от теста	Положителен	1	Вярно (действително) положителен резултат (TP)	Невярно (фалшиво) положителен резултат (FP)	Прецизност (положителна предсказана стойност, PPV) PPV = TP/(TP+FP)
	Отрицателен	0	Невярно (фалшиво) отрицателен резултат (FN)	Вярно (действително) отрицателен резултат (TN)	Отрицателна предсказана стойност (NPV) NPV = TN/(TN+FN)
Коефициент погрешно класифициране MR = (FP+FN)/Total			Чувствителност (SE или TPR) SE = TP/(TP+FN)	Специфичност , (SPC или TNR) SPC = TN/(FP+TN)	Точност ACC + MR = 1 ACC = (TP+TN)/Total

Едно от съществените предимства, при използване на матрицата на объркване като инструмент за оценяване, е това, че позволява по-подробен анализ, отколкото дава простото отношение на правилно класифицираните примери, т.е. *точността*, която може да предостави подвеждащи резултати, ако наборът от данни не е балансиран, т.е. когато има съществени разлики в броя между четирите групи от резултати. В общия случай, матрицата е силно небалансирана. Самата матрица на объркването е сравнително проста за разбиране, но свързаната с нея терминология *може да бъде твърде объркваща*.

Точност

Точността (Accuracy, ACC) на класификация се определя от общия брой вярно положителни и вярно отрицателни резултати (TP+TN), разделен на общия брой на обхванатите от теста лица (TP+TN+FP+FN), Табл. 1.

Точността отделя всички, които тестът вярно е определил като болни и като здрави, от всички изследвани в случайната извадка. Това може да звучи като добър базов показател за оценяване на ефективността на теста, но в една неуравновесена извадка, това не е толкова сигурно. Като мяра за ефективността на теста, показателят точност е **неподходящ** при оценяване на небалансирани класификации. Така, например, нека в една силно небалансирана извадка, състояща се от 95 верни отрицателни и 5 верни положителни резултати, класифицираме всички резултати като отрицателни, т.е. TN = 95, FN = 5, при което въпреки това точността остава висока: ACC = 0,95, но 5-ма болни от COVID-19 остават неоткрити!

Втори пример, нека в друга небалансирана извадка болните са 20, а здравите – 80 души, резултатите от класифицирането обаче са: TP = 19; TN = 80; FN = 1 и FP = 0. Тогава точността ще бъде още по висока ACC = 0,99, но един болен все пак няма да бъде открит, т.е. въпреки привидно голямата точност, тестът не е открил един болен човек, който ще продължи да заразява.

Твърде лесно, тази класификация постигна голяма точност (0,99 или 99 %), но не открива всички заразени (болни). Открива само 95 % (19/20) от болните. *Запомнете*, точността, като инструмент за оценяване, показва едно единствено нещо, **колко души от извадката имат верен (коректен) резултат**, и нищо друго!

Нормата на погрешно класифициране (Misclassification Rate, MR), отделя тези изследвани, които са получили невярно положителен и невярно отрицателен резултат (FP+FN), от всички получени от теста резултати (TP+FP+TN+FN), Табл. 2. Тя се определя чрез точността по следния начин: $MR = (1 - ACC)$. Тя, също така, не може самостоятелно да бъде използвана като инструмент за оценяване.

Нека да продължим с примера, ако точността ACC, като показател, не може да бъде използвана за разкриване на ефективността на едно тестване (изследване, оценяване, класифициране), то тогава трябва да се потърсят нови показатели, които да проявят неговата ефективност.

Прецизност

Прецизността, или *положително предсказаната норма (Precision; Positive Predictive Value, True Positive Rate, PPV)*, отделя тези, които са получили верен положителен резултат (TP), от всички изследвани от теста с положителен резултат (TP+FP), Табл. 1. Прецизността позволява да се отговори на въпроса, колко от тези, получили положителен резултат от тестването (TP+FP), действително са болни от COVID-19, т.е. те са получили верен положителен резултат (TP). Или казано по друг начин, *предсказвайки „да“, колко често това е вярно?*

Нека разгледаме, една случайна извадка с 180 положителни резултата, които принадлежат на минорния клас (болни): TP+FP = 135 + 45 = 180, при което прецизността ще бъде $PPV = TP/(TP+FP) = 0,75$. Тази прецизност е приемлива, но не е нещо изключително. Нека сега вземем същия набор от данни, при което само 40 положителни резултата принадлежат на минорния клас: TP+FP = 36 + 4 = 40, при което прецизността ще бъде $PPV = 0,90$. В този случай, въпреки че още по-малко резултати принадлежат към минорния клас на положителните резултати, то прецизността е значително по-голяма, т.е. неверните положителни резултати са съществено по-малко. *Колкото прецизността е по-висока, толкова по-малко ще бъдат неверните положителни резултати!* Колкото по-ниска е прецизността, толкова повече от изследваните ще бъдат поставени в

положението да бъдат обявени за болни от COVID-19. *Представете си*, каква отговорност се поема към здравите хора, които са обявени неоснователно за болни, към техните семейства и техните близки?!

Чувствителност

Чувствителността (Sensitivity, Power или Recall, SEN), или *вярно положителната норма* (True Positive Rate, TPR), като показател, отделя всички с верен положителен резултат (TP) от всички, които са болни от COVID-19 (TP+FN), Табл. 1. Този показател е твърде важен, тъй като отговаря на въпроса, колко от действително болните от COVID-19 са тествани (вярно) положително. Този показател дава отрицателен резултат за здравите. Въпросът, който трябва да се зададе на това място е: „*Бихме ли искали болни хора да си мислят, че не са болни в действителност?*“. Каква е отговорността да кажеш на някой действително болен от COVID-19, че не е? Отговорност пред него, пред близките му и пред обществото!

Много често, **чувствителността се противопоставя на прецизността**, тъй като, при еднакъв брой верни положителни резултати, колкото чувствителността е по-ниска, толкова повече действително болни хора се обявяват за здрави, и обратното, колкото прецизността е по-ниска, толкова повече здрави хора се обявяват за болни, т.е. грешките от втори и първи тип доминират в реализираната класификация, Табл. 1.

Чувствителността, обаче, е по-полезна за лекаря, докато прецизността е по-полезна за пациента. *Прецизността* показва на изследвания шансовете да има заболяване, ако има положителен резултат. *Чувствителността* показва на лекаря шансовете да идентифицира болния, ако има положителен резултат. Тя е делът на действително положителните, които действително се тестват като положителни, т.е. определя доколко един тест е в състояние да открие действително положителни индивиди в една популация. Високо чувствителният тест означава, че има малко фалшиво отрицателни резултати и по този начин се пропускат по-малко случаи на заболяване.

Съществува още един показател, който се използва за оценяване едновременно на чувствителността и на прецизността. Това е *средната хармонична стойност* от стойностите на прецизността и на чувствителността, наречена **F-мярка** (F-measure) или балансиран **F-резултат** (F-score), която представлява отношението на средно квадратичната стойност (PPV.SEN) към средно аритметичната стойност (PPV+SEN):

$$F\text{-measure} = 2(PPV.SEN)/(PPV+SEN) \quad (1)$$

Специфичност

Специфичността (Specificity, Selectivity, SPC) отделя всички изследвани, които тестът е предвидил вярно, че са здрави (TN), от всички, които действително са здрави (FP+TN). Специфичността дава отрицателен резултат за заболялите от COVID-19. Тя се разглежда още като вярно отрицателна норма (True Negative Rate, TNR). Тя помага да се отговори на въпроса, колко вярно тестът предвижда, че изследваните не са болни от COVID-19? Въпросът, който вие трябва да си зададете е следният: „*Бихте ли искали да се поставите под карантина поради грешен тест?*“.

Нека, тест с 95 % специфичност бъде приложен върху случайна извадка от 100 души, при което той произвежда 5 невярно (фалшиво) положителни резултати. Ако реалната честота на заразяване е 1 %, това означава, че само един от 100 би бил действително заразен. Такъв тест би открил една действителна инфекция (или един болен) и пет „невярно положителни“ инфекции. Подобно неточно тестване може да дезинформира лицата, които вземат решения, включително за такива критични показатели като смъртността от COVID-19. Още нещо, повечето „невярно положителни“ тестове могат драматично да изкривят тези резултати, предизвиквайки паника и страх.

Това може да наложи една погрешна теза – за много ниската смъртност от COVID-19, защото „има много случаи на заболяване, но малко хора са умрели“. Така може да се произведе у хората фалшивото чувство за сигурност!

Много често в практиката се използва допълнителния показател на специфичността, *невярната положителна норма* (False Positive Rate, FPR), който отделя всички изследвани, които тестът е обявил невярно за болни (FP), от тези, които действително са здрави (FP+TN):

$$FPR = (1 - SPC) = 1 - TN/(TN+FP) = FP/(FP+TN), \quad (2)$$

Специфичност и чувствителност

Специфичността определя дяла на изследваните без заболяване, които се тестват отрицателно, т.е. правилно са изключени (квалифицирани) от теста. В същото време, чувствителността определя дела на изследваните с болестта, които се тестват положително, т.е. правилно са идентифицирани от теста.

С повишаването на специфичността, чувствителността намалява и обратно, т.е. двата базови показателя се намират в обратно пропорционална зависимост. *Чувствителността* е процентът на верните положителни резултати, напр. чувствителност $SEN = 0,85$ означава, че 85 % от положителните резултати са верни (действителни, истински). *Специфичността* е процентът на верните отрицателни резултати, напр. специфичност $SPC = 0,85$, означава че 85 % от отрицателните резултати са верни. Тест, който има специфичност $SPC = 1,00$ или 100 %, ще идентифицира 100 % от пациентите, които нямат заболяването. Тест, който има специфичност $SPC = 0,85$ или 85 %, ще идентифицира 85 % от пациентите, които нямат заболяването. *Чувствителността и специфичността се явяват основни показатели на всеки бинарен небалансиран тест.* Клинично, тези понятия са важни за потвърждаване или изключване на заболяването по време на скрининг.

Така, например, нека бинарен тест открива в една случайна извадка, състояща се от 50 болни от COVID-19 и 50 здрави, 50 вярно положителни резултати (TP) и 50 невярно отрицателни резултати (FN). Това означава, че тестът е 100 % *чувствителен* и 100 % *специфичен* – това е идеалният тест, който класифицира напълно вярно, и болни, и здрави. Ползата е и за лекаря, който трябва да лекува болните, и за популацията, защото здравите ще знаят, че са действително здрави.

В началото на април 2020, беше съобщено (*John Gage, Washington Examiner, 05 април*) с тревога, че едва 70 % от всички действително болни от COVID-19 (SARS-CoV-2 инфекция) се откриват чрез широко използваните тестове. Това означава, че чувствителността на използваните тестове е едва $SEN = 0,70$ или 70 %. Нека бъде разгледан именно този случай, но при голяма специфичност на теста $SPC = 0,95$, т.е. това е тест, който би трябвало да идентифицира преди всичко истински здравите. Неверните положителни резултати (FP) би трябвало да бъдат много малко. Изследването се прави при непрекъснато нарастване на броя на действително болните от COVID-19 в изследваната извадка, Табл. 3.

Очевидно става това, че този тест трудно би работил в полза на лекаря, тъй като ниската чувствителност на теста произвежда твърде много неверни отрицателни резултати, т.е. болни се класифицират като здрави. При това абсолютният брой на тези неразкрити болни нараства с увеличаването на болните в извадката. Действително, по-малко здрави ще бъдат обявявани за болни (FP) и това е добре. Тревогата идва, обаче, от това, че все повече болни ще бъдат пропуснати при тестването (FN) и те ще продължат все по-интензивно да заразяват популацията. За да се излезе от това състояние, трябва да се използват тестове с висока чувствителност, които действително работят в полза на лекаря и на болните от COVID-19.

Таблица 3.

Брой на изследваните, които имат действително Covid-19 или „вероятност преди теста“, да	Верен положителен резултат (TP), бр.	Неверен отрицателен резултат или грешка от II тип (FN), бр.	Чувствителност (SE = TP/(TP+FN), -	Брой на изследваните, които действително нямат Covid-19, не	Неверен положителен резултат или грешка от I тип (FP), бр.	Верен отрицателен резултат (TN), бр.	Специфичност (SPC = TN/(TN+FP), -
10	7	3	0,70	90	4	86	0,95
20	14	6	0,70	80	4	76	0,95
30	21	9	0,70	70	3	67	0,95
40	28	12	0,70	60	3	57	0,95
50	35	15	0,70	50	2	48	0,95
60	42	18	0,70	40	2	38	0,95
70	49	21	0,70	30	1	29	0,95
80	56	24	0,70	20	1	19	0,95
90	63	27	0,70	10	0	10	0,95

Този пример поставя един основателен въпрос: *Трябва ли тестът за скрининг да е чувствителен или специфичен?* Валидността на теста е неговата способност за точно идентифициране и на болните, и на здравите лица. Идеалният тест за скрининг трябва да бъде изключително чувствителен, т.е. с висока вероятност за откриване на заболяването, и изключително специфичен, т.е. с голяма вероятност за откриване на здравите – на тези, които не са болни.

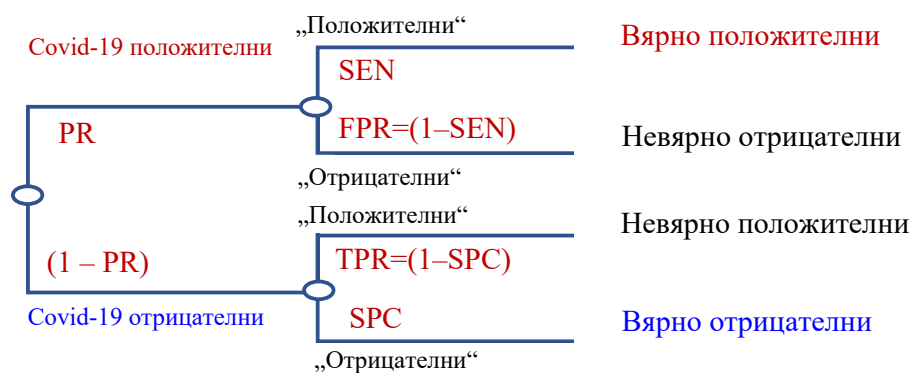
Преобладаването или „широкото разпространение“ (*Prevalence*, PR) отделя болните от COVID-19 (TP+FN), като дял от всички изследвани от популацията (TP+TN+FP+FN), Фиг. 1 и 3:

$$PR = (TP+FN)/(TP+TN+FP+FN). \quad (3)$$

AUC–ROC подход за оценяване на ефективността на теста

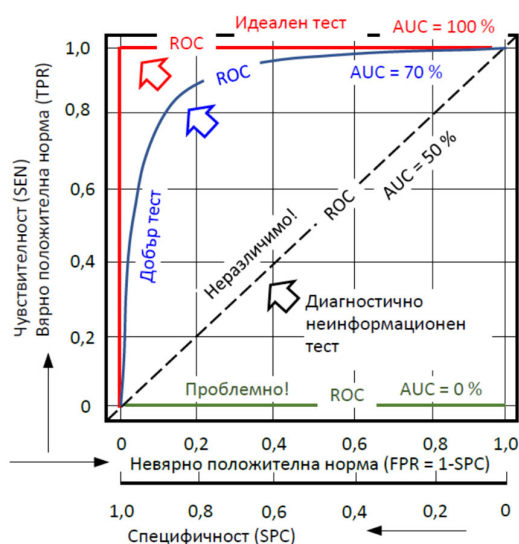
Чувствителността SEN и специфичността SPC позволяват да бъдат отделени напълно „положителните“ – „вярно положителните“ и „невярно положителните“, т.е. болните от COVID-19, като се използва чувствителността SEN и невярно положителната норма FPR = (1 – SPC). По същия начин, могат да бъдат отделени напълно „отрицателните“ – „вярно отрицателните“ и „невярно положителните“, т.е. здравите, като се използва специфичността SPC и невярно отрицателната норма FNR = (1 – SEN), фиг. 3.

Този подход може да позволи валидирането на тестове за COVID-19, като се използват по същество двете базови характеристики на теста – неговата чувствителност и специфичност, фиг. 3.



Фигура 3.

Ефективността на класификацията, или на използвания тест, може да се определи чрез използването на т.нар. *настройка на праговете*. Когато става дума за решаване на тази съществена задача, както и да се визуализира ефективността, обикновена се използва подхода AUC-ROC, където ROC (Receiver Operating Characteristic) е кривата на вероятностите, а AUC (Area Under The Curve) е площта под кривата на вероятностите и е показател или мяра за разделянето на двата класа. По-голямата площ AUC, разкрива възможностите на по-добрия модел да разграничава по-добре изследваните от популацията със заболяването и тези без заболяването. Тя определя количествено способността на избрания модел на класификация да прави разлика между двата класа. Много често тя се изписва като AUROC, т.е. зона под експлоатационната характеристика на приемника. Наименованието на ROC-кривите е наистина странно – то означава *оперативна характеристика на приемника*, като в действителност то отразява исторически първоначалното им използване – за отделянето на полезния *сонарен* сигнал (напр., от подводница) от т.нар. „шум“, предизвикан от преминаващ рибен пасаж или отделянето на полезния *радарен* сигнал (напр., от самолет) от шума, предизвикан от ято птици.

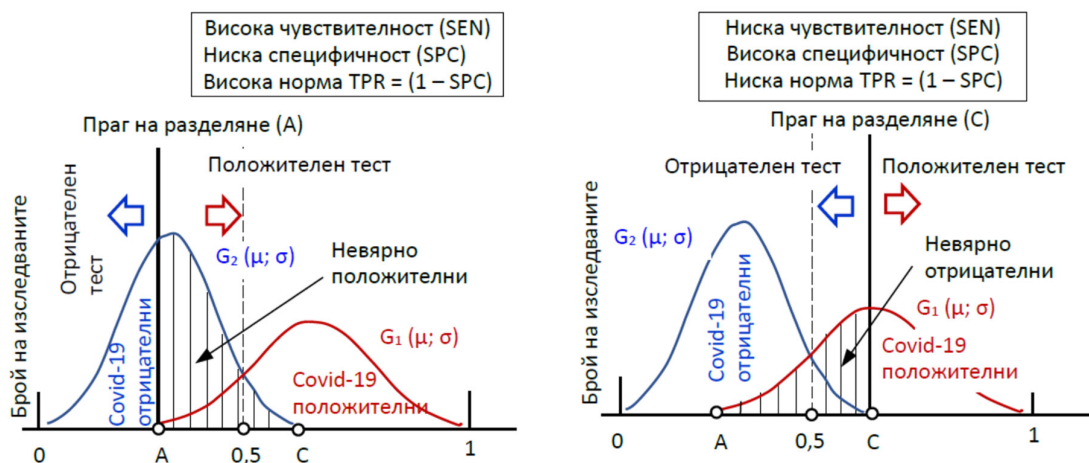


Фигура 4

ROC-кривата в графичен вид изразява зависимостта между чувствителността (SEN) или невярно положителна норма – $FPR = (1 - SPC)$, където SPC е специфичността на теста. Тъй като, чувствителността по същество е вярно положителната норма (TPR), то зависимостта е между *вярно положителна норма* (TPR) и *невярно положителна норма* (FPR). Идеалният модел ($AUC = 100\%$), т.е. идеалната класификация, би имала: максимална чувствителност (SEN) при максимална специфичност (SPC) или минимална невярна положителна норма (FPR). Лошият модел е с $AUC \approx 0$. За да бъде приет един тест за диагностично информативен, то трябва $AUC > 50\%$. При $AUC = 50\%$, тестът на практика е напълно безполезен, фиг. 4.

ROC-кривите са невероятно полезен инструмент за оценяване и сравняване на прогнозни модели. Те се използват често, за да покажат връзката, или по-точно компромиса, който се прави, между клиничната чувствителност и специфичност, за всеки един тест, с цел възможното му изключване от употреба или за един обоснован избор измежду множество налични тестове. При съвременното им използване, ROC-кривите са един добър инструмент за оценяване на възможностите на прогнозиращия модел да различи действително положителните и отрицателните в една случайна извадка (популация). Правилото е, че моделът трябва да прогнозира вярно не само положителното като положително, но и отрицателното като отрицателно, или здравите като здрави и болните като болни.

Когато се намали прага на разделяне (A), се получават повече положителни стойности, като по този начин се увеличава чувствителността и се намалява специфичността. При това положение, рискуваме да получим невярно положителни резултати, т.е. здрави хора да бъдат обявени за болни от COVID-19. Този подход, с максимално намаляване на прага на разделяне, привидно е в полза на лекаря, тъй като му позволява да открие и лекува всички заболели, фиг. 5. Този подход, обаче, косвено търси също така максимална полза за обществото чрез проследяване и ограничаване на заболяването. Интересите на обществото и на лекаря не може да се противопоставят. Лекарят, медицината и фармацевцията служат на обществените интереси. При отсъствието на лекарства и ваксина срещу COVID-19, обаче, той остава единственият смислен подход, който трябва да бъде поддържан.



Фигура 5

Другият небалансиран подход се свързва с увеличаване на прага на разделяне (C), като при това се получават повече отрицателни стойности, и по този начин се увеличава специфичността и се намалява чувствителността на теста. При този подход, рискуваме да получим много невярно отрицателни резултати, т.е. действително болни от COVID-19 да бъдат обявени за здрави, което носи риска от разпространяване на пандемията. Този подход, с максимално увеличаване на прага на разделяне, е в полза на обществото и неговите дейности, тъй като му позволява да открие и ползва всички здрави (и преболедували) хора, фиг. 5. Всъщност, този подход търси решение в полза на политиците и на бизнеса – той трябва да окаже максимално положително влияние върху работната сила, търговията, отдиха и пътуванията, финансовите пазари, управлението и сигурността на държавата.

Естествено е това, че съществува напълно балансиран подход (вероятност 0,5), при което се произвеждат в еднакво количество невярно положителни и невярно отрицателни резултати от теста. Всъщност, тук се проявява ролята на верния избор чрез определяне на прага на разделяне – нисък, висок или балансиран, това обаче вече е обект на здравна политика.

Вместо заключение

Непрекъснато чуваме все по-високо извисяващи се гласове, в които долавяме паника и страх: „Тестовите (за антитела) трябваше да помогнат за максимално бързото възстановяване на най-пострадалите от заболяването COVID-19 страни – САЩ, Великобритания, Испания, Италия, Франция, и т.н. Вместо това те внесоха повече объркване отколкото яснота на фона на бързо разрастващата се пандемия“. На това, СЗО отговори настоятелно: „Тествайте, тествайте, тествайте..!“.

Можем ли да добавим нещо повече, след като поставихме и разгледахме основните положения, свързани с базовите характеристики и ефективността на тестовите, определена от целенасочения избор на прага на изместване?

Да, на първо място, валидирайте и сертифицирайте (за безопасност) всички тестове, определете т. нар. „златен стандарт“ и тогава тествайте, тествайте, и пак тествайте. На второ място, поддържайте качеството на работа в медико-диагностичната дейност, като използвайте добрите международни практики – действително, акредитирана ли е вашата лаборатория по международния стандарт ISO 15 189:2012. На трето място е потвърдителното тестване: СЗО препоръчва на страните, чиито национални лаборатории са с ограничен опит в тестването на заболяването COVID-19, да изпратят първите си пет положителни и първите си десет отрицателни проби за COVID-19, в една от 16-те референтни лаборатории на СЗО за потвърдителни тестове. Референтните лаборатории са разпределени териториално така: в Азия са 7, в Европа – 5, в Африка – 2, в Северна Америка – 1 и в Австралия – 1.

И все пак, трябва да повторим, като допълним казаното от *Colin West*: „Трябва да спрем да се преструваме, че медико-диагностичната дейност и тестовите за COVID-19 са перфектни!“. Броят на случаите на COVID-19, потвърдени чрез тестване, е значително по-малък от броя на регистрираните инфекции, казват служители на общественото здравеопазване. Въпреки всичко, броят на тестовите, които могат да бъдат направени, също така остава ограничен от доставките дори на такива консумативи, като тампони и реактиви, както и от недостига на обучени хора и защитни средства. Ето защо, отново трябва да напомним: „Точно заради това, не спирайте да миете ръцете си, не спирайте да носите маска и не пренебрегвайте физическото дистанциране!“.

Когато се говори за ефективността на един тест, винаги трябва да се поставяме на мястото на лекаря, който трябва да представи резултата от изследването пред поредния

пациент. Ако резултатът е положителен, то колко вероятно е този пациент наистина да е заболял от болестта COVID-19. *Колко трябва да се притеснява този лекар, съобщавайки тази вест?* – А ако действително не е заболял? Ако тестът е отрицателен, колко вероятно е този пациент наистина да не е заболял от болестта COVID-19. *Колко успокоен трябва да бъде този лекар, съобщавайки тази вест – а ако е заболял?* Когато не познаваме добре вируса и самото заболяване, когато няма лекарства и ваксини, когато се обучаваме в хода на безпрецедентна пандемия, какви тестове ни трябва, за да не се притесняваме и да не се успокояваме излишно? Отговорът оставяме на вас, читателю. Надяваме се, че след като стигнахте дотук, вече знаете и можете да направите своя информиран избор.