

# SEED Hematologia



## Wstęp do statystyki medycznej

### Znaczenie badań statystycznych

Nowoczesna medycyna często opiera się na analizie statystycznej częstości zachorowań, analizie prawdopodobieństwa powodzenia danej metody terapii, poszukiwaniu wartości „prawidłowych” w badaniach laboratoryjnych, etc. Ma to znaczenie dla twierdzeń, które zostały oparte na tych statystykach. Znaczenie wyników badania lub uzyskanych wartości laboratoryjnych jest także zależne od warunków statystycznych, które opisują badanie lub definiują limity określone dla podejmowanych decyzji. Znajomość zasad statystycznych jest dlatego istotnym narzędziem w interpretacji wyników laboratoryjnych. Co to oznacza dla pacjenta gdy np. dana wartość jest mierzona.

### Pobieranie materiału i związany z nim zakres odchyień

Teoretycznie pomiar diagnostyczny jest wykonany z zamiarem nabycia wiedzy na temat stanu pacjenta (stężenie lub zmiana, czas reakcji, morfologia tkanek etc.). W wielu wypadkach do tego celu pobierana jest próbka, na przykład moczu lub krwi. Porcja każdej z tych próbek jest następnie wykorzystywana do różnych badań. Zakładając że określony składnik, na przykład stężenie krwinek białych nie jest wszędzie identyczne, ale jest rozłożone przypadkowo to można by przyjąć, że właściwie otrzymujemy przybliżony stan pacjenta. Jeśli mocz pacjenta zawiera średnio 10 WBC/ $\mu$ l,

to całkowicie możliwe jest, że w niektórych jego częściach może znajdować się np. 12 WBC/ $\mu$ l. Może to być równoważone przez inne miejsce, przez odpowiednio niższe stężenie. Dlatego więc, statystycznie rzecz ujmując, powoduje to potencjalny zakres wahania związany z pobieraniem próbki. Jest to określone przez samo stężenie. W wypadku małego stężenia analitu, nawet niewielkie wahania w wartościach bezwzględnych mogą skutkować drastyczną zmianą wartości odsetkowych. W przypadku opisanym wyżej, statystyczna zmienność powoduje błąd 20%!

Inaczej ma się sprawa w przypadku wahań w wartościach bezwzględnych pomiaru WBC w prawidłowej próbce krwi. Mogą one nie wywoływać większych zmian, a jest to spowodowane przez znacznie wyższe stężenie komórek (od ok. 4 – 10 x 10<sup>3</sup>/ $\mu$ l).

### Współczynnik zmienności

Powyższy przykład opisuje wahania statystyczne, które są niezależne od użytej metody. Spowodowane są usunięciem części próbki z całości (tj. z ogólnie pojętego sytemu krwionośnego pacjenta lub pewnej części z próbki) i losowego rozkładu danego elementu w całym układzie. Warunkuje to statystyczny współczynnik zmienności (CV), który określa zakres wahania wokół wartości „prawdziwej” w badanym środowisku i jest całkowicie zależny od stężenia.

Jest to przedstawione w tabeli Rümke (tab. 1), pokazującej jak wiele komórek należy zbadać aby otrzymać pewną wiarygodność zmierzonej wartości. Jeśli pomiary w laboratorium wskazują mniejszy współczynnik zmienności należy pamiętać, że te ostatnie odnoszą się do średniej z pomiarów, a nie „prawdziwej” wartości. (patrz ryc. 1).

Oprócz współczynnika zmienności wyróżnia się inne czynniki wpływające na powstawanie różnic np. zmienność pipetowanej objętości, subiektywna ocena, możliwy brak standaryzacji (np. przez zmieniający się personel), itp. W tym wypadku, pomiar parametru zawsze odbywa się przy istniejącym konflikcie między precyzją a praktycznością.

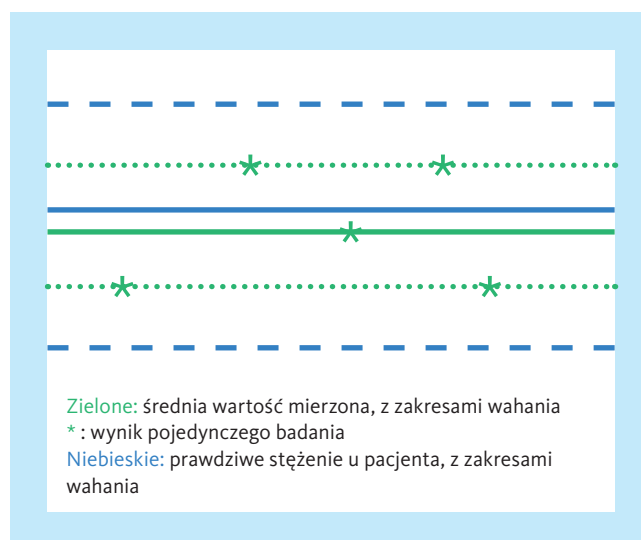
Pytanie dotyczące użyteczności klinicznej ma znaczenie w podejmowaniu decyzji – jak dokładnie należy oznaczyć parametr w celu podjęcia odpowiedzialnej decyzji klinicznej?

W tym wypadku na dwa sposoby może być pomocna automatyzacja: w wysokich stężeniach (np. RBC w krwi pełnej, około  $5 \times 10^6 / \mu\text{l}$ ), CV jest niewielkie dzięki zliczaniu bardzo dużych ilości komórek. Jednocześnie minimalizowane są inne źródła błędów, ponieważ pobranie i ocena próbki odbywają się zawsze w ten sam sposób.

## Współczynnik zmienności (CV)

Współczynnik zmienności wskazuje niepewność pomiarową w oznaczonym stężeniu, w odniesieniu do tego stężenia. Jest miarą nieprecyzyjności metody i obliczany jest jako iloraz odchylenia standardowego i średniej. Wynik jest przedstawiony w procentach. W przeciwieństwie do odchylenia standardowego, które przedstawia wartości bezwzględne zakresu wahań, CV umożliwia pokazanie jak znaczny jest wpływ nieprecyzyjności przy danym stężeniu.

Przykład: Odchylenie standardowe wynoszące 2 przy stężeniu 2 komórki/ $\mu\text{l}$  oznacza wysoki współczynnik zmienności na poziomie 100%. Jednakże, takie samo odchylenie standardowe przy stężeniu 200 komórek/ $\mu\text{l}$  jest znacznie mniej problematyczne, a CV wynosi tu tylko 1%.



**Ryc. 1** Wartość prawdziwa (gruba niebieska linia) ze statystycznym współczynnikiem zmienności (niebieska linia przerywana). Średnia z badania (linia zielona) z zakresami wahań mierzonej wartości (zielona linia przerywana). CV badania wydaje się tu mieć mniejszy zakres niż możliwy statystycznie. Jednakże pomiary wartości tylko przypadkowo znajdują się w węższym zakresie wokół średniej. Są one rozrzucone w całym zakresie współczynnika zmienności. Dalsze pomiary pokazałyby statystyczne CV i średnią bliższą wartości prawdziwej.

## Przedziały ufności

Kiedy już uzyskamy wartość danego parametru, następuje pytanie, jakie informacje ona ze sobą niesie. W celu odpowiedzi czy dana wartość jest patologiczna, ustalono zakresy wartości referencyjnych, które określają rozkład danej cechy wśród zdrowej populacji. W celu określenia zakresów referencyjnych, zbierane są wartości wyników zmierzonych u zdrowych ludzi (czasem także od chorych, jeśli dane zaburzenie nie wpływało na wartość wyniku). Z tych wartości wyznacza się 95% przedział ufności, czyli zakres, w którym mieści się 95% zmierzonych wartości. Przedział ten zazwyczaj stosowany jest jako zakres referencyjny.

Dlatego należy pamiętać, że wynik poza tym zakresem niekoniecznie oznacza, że badana osoba jest chora. Raczej można powiedzieć, że 5% populacji, ma wartości niższe lub wyższe od zakresów referencyjnych. Dlatego zachodzi pytanie: czy badany należy do tej grupy? I odwrotnie, zmierzona wartość w obrębie zakresów referencyjnych nie zawsze oznacza, że pacjent jest zdrowy.

**Tabela 1** Wycinek z tabeli Rümke przedstawiający 95% współczynniki zmienności dla podanych całkowitych liczb zliczonych komórek.  $a$  = znaleziony procentowy wskaźnik populacji;  $n$  = liczba zliczonych komórek;  $X-Y$  = 95% przedział ufności wyniku. Oznacza to, że wynik będzie się mieścić w tym zakresie w 95% przypadków, poniżej w 2,5% przypadków oraz powyżej w 2,5% przypadków. Na przykładzie manualnego różnicowania 100 WBC, oznacza to, że prawdziwa wartość w 95% zliczeń może się mieścić w zakresie 0-9 w przypadku znalezienia 3% eozynofili. Nawet przy policzeniu 1000 WBC nie można wykluczyć, że obecnych jest raczej 5% niż 3% eozynofili. Tabela może być rozbudowana odpowiednio dla większych wartości procentowych (np. odpowiadającym neutrofilom), ale również dla większej liczby zliczonych komórek (jak np. w analizatorach automatycznych).

a	n = 100	n = 500	n = 1000
1	0-4	0-1	0-1
2	0-6	0-3	0-2
3	0-9	1-5	2-5
4	1-10	2-7	2-6
5	1-12	3-8	3-7
6	2-13	4-9	4-8
7	2-14	4-10	5-9
8	3-16	5-11	6-10
9	4-17	6-12	7-11
10	4-18	7-13	8-13

## Porównanie testów i odniesienie do metody referencyjnej

Rzeczywiste pytanie diagnostyczne jest redukowane do wyrażenia przeciwnych „chory – zdrowy” lub „terapia przynosi efekty – terapia nie przynosi efektów”. Decyzja ta jest oparta na pewnych, tak zwanych wartościach granicznych (ang. cut-off) lub przyjętych limitach. Określenie jakości testu, odbywa się przez porównywanie z przyjętymi badaniami, na przykład ze „złotym standardem” (czyli rzekomo najlepszą dostępną metodą) lub z metodą referencyjną, czyli metodą która została oficjalnie uznana jako standard w określaniu danych parametrów.

**Tabela 2** Przykład „tabeli czterokrotnej” dla ogólnego porównania metod

	Stary test dodatni	Stary test ujemny
Nowy test dodatni	A	B
Nowy test ujemny	C	D

Wiele testów może być porównanych na podstawie „tabeli czterokrotnej”. W ten sposób tak zwana „zgodność” oznaczająca zbieżność dwóch wyników badań, może być wykorzystana do wstępnego porównania metod. W tym procesie należy wziąć pod uwagę, że wynik starej metody niekoniecznie jest prawidłowy.

Sytuacja wygląda inaczej jeśli wyniki są porównywane z metodą referencyjną. Wynik jest wtedy poprawny z definicji (co może stanowić problem jeśli metoda referencyjna nie jest bardzo precyzyjna).

## Metoda referencyjna

Metoda referencyjna jest metodą analityczną uznawaną przez określone stowarzyszenia zawodowe, jako metoda najbardziej wiarygodna. Oznacza to, że wartość uzyskana tą metodą jest traktowana jako „prawdziwa”. Jednakże, termin ten często jest używany niepoprawnie, jako określenie najpowszechniej używanej metody. Może to powodować problemy ponieważ nawet powszechnie używana metoda, może dostarczać fałszywych wyników, nie tak łatwo określanych jako „prawdziwe”. Ma to wpływ zarówno na wnioski przy porównaniu metod, jak i w określeniu metody do stosowania. W tym wypadku lepiej jest użyć określenia „wybrana metoda porównawcza”.

## Czułość i swoistość

Użycie „tabeli czterokrotnej” pozwala obliczyć jakie jest prawdopodobieństwo, że pozytywny wynik testu jest prawdziwie dodatni, jak i prawdopodobieństwo, że negatywny wynik testu jest prawdziwie ujemny. Krótko mówiąc: jak często wynik testu będzie poprawny. Tabela 3 przedstawia przykład oceny testu wykonanego tą właśnie metodą.

W tym przypadku próbki niezgodne z wymaganiami są fałszywie dodatnie (B) lub fałszywie ujemne (C). Teraz za pomocą porównania tych danych z metodą referencyjną można otrzymać następujące informacje:

- Jak często jest wykrywana próbka prawdziwie dodatnia? (czułość)
- Jak często jest wykrywana próbka prawdziwie ujemna? (swoistość)

Te dwie charakterystyki są wobec siebie przeciwstawne: im bardziej test jest czuły, tym wyższe jest prawdopodobieństwo fałszywego określenia próbki negatywnej jako pozytywnej. Podobnie ma się sytuacja w przypadku testów zbyt swoistych, które mogą mieć tendencję do fałszywego określania próbki jako negatywna. Istotny jest zatem obszar zastosowania badania.

### Czułość

Czułość określa jak często próbka dodatnia jest określana przez badanie jako pozytywna. W przypadku niskiej czułości testu pomija on wiele dodatnich próbek.

### Swoistość

Swoistość określa jak często ujemna próbka jest określana przez badanie jako negatywna. Niska swoistość powoduje wiele fałszywych ostrzeżeń.

Możliwa jest optymalizacja kluczowych cech testu, np. dopasowanie wartości cut-off. W zależności od specyficznego zapotrzebowania, pożądanym kryterium może być wysoka czułość lub swoistość. Nie zawsze konieczne jest aby znaleźć optymalną równowagę między tymi dwoma charakterystykami. Co więcej, czułość i swoistość niewiele mogą powiedzieć o konkretnym przypadku. Przedstawiają raczej ogólne informacje dotyczące wiarygodności testu.

### Wartości predykcyjne

Z tych danych mogą być obliczone tak zwane „wartości predykcyjne”. Tam gdzie czułość i swoistość pozostawiają pytanie jak często stan pacjenta jest poprawnie rozpoznany, dodatnia i ujemna wartość predykcyjna wskazuje jak często dodatnie lub ujemne wartości badań są rzeczywiście poprawne.

Wysoka dodatnia wartość predykcyjna pozwala bezpiecznie stwierdzić, że próbka dodatnia towarzyszy chorobie. Jednakże, ujemny wynik badania nie koniecznie oznacza „brak choroby”. Tym samym stan chorobowy może być przeoczony. Analogicznie, wysoka ujemna wartość predykcyjna umożliwia zaklasyfikowanie ujemnej próbki jako niepodważalnej.

### Dodatnia wartość predykcyjna

Dodatnia wartość predykcyjna wskazuje na prawdopodobieństwo obecności zjawiska, np. pewnej choroby, jeśli badanie na nią daje wynik dodatni.

### Ujemna wartość predykcyjna

Ujemna wartość predykcyjna wskazuje na prawdopodobieństwo, w przypadku uzyskania ujemnego wyniku testu, że badana osoba rzeczywiście nie cierpi na chorobę, na którą jest badana.

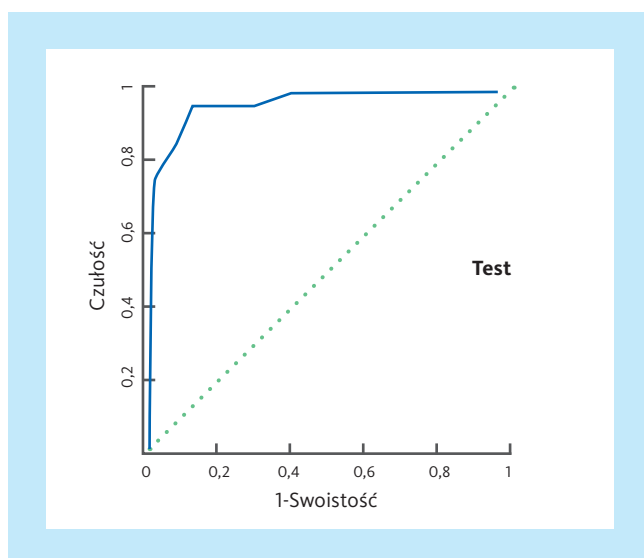
Znaczenie ma tu także częstość (rozpowszechnienie) występowania danej choroby. Jeśli choroba jest bardzo rzadka, obserwowany będzie większy odsetek próbek negatywnych wśród całej populacji. Bardzo ciężko jest otrzymać dobrą dodatnią wartość predykcyjną. Wartości predykcyjne dobrze opisują punkt widzenia lekarza lub pacjenta. Takie osoby chcą wiedzieć co znaczy wynik ujemny albo dodatni: „Jak duże jest prawdopodobieństwo, że cierpię na chorobę pomimo ujemnego wyniku?” lub „Jak bardzo powinienem się martwić o wynik dodatni?”. Dla niektórych testów przesiewowych akceptowalne są wysokie fałszywie dodatnie wartości, ponieważ testy takie są szybkie i tanie. Wynik dodatni natomiast może być potwierdzony dodatkowym badaniem o wyższej czułości i swoistości przeprowadzonym z udziałem mniejszej liczby podmiotów.

Biorąc pod uwagę rozpowszechnienie, należy rozważyć rodzaj pobieranej próbki. Częstotliwość występowania danej jednostki chorobowej w całej populacji może rzadko odpowiadać częstości występowania jej w próbkach w laboratorium szpitalnym. Większość pacjentów jest w szpitalu z pewnego powodu; dlatego więc prawdopodobieństwo, że cierpią oni na chorobę jest znacznie wyższe niż wśród ogólnej populacji.

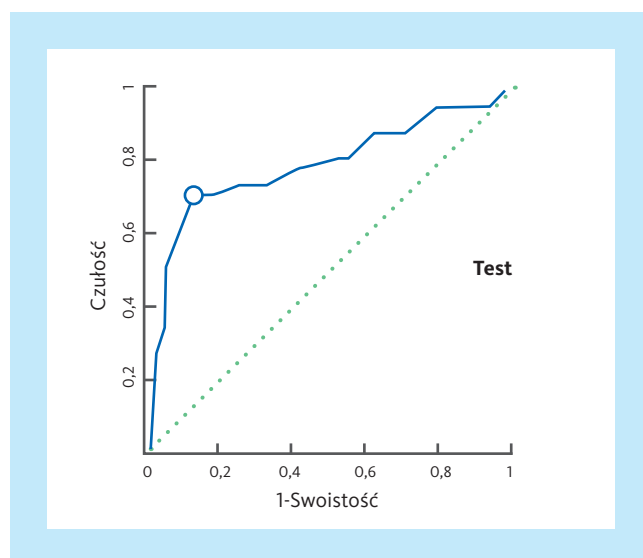
Częstość występowania wskazuje także na prawdopodobieństwo, że dana choroba jest obecna, nawet jeśli nie ma jakichkolwiek innych danych o pacjencie. Dobry test diagnostyczny powinien zwiększać to prawdopodobieństwo w sposób klarowny: test dodatni powinien dawać znacznie wyższą pewność, że choroba jest obecna. Jeśli tak nie jest, to test nie przyczynia się do zdobywania nowych informacji.

**Tabela 3** Przykład „tabeli czterokrotnej” użytej do określania jakości testu diagnostycznego. Przedstawione dane pokazują czułość 85% (67 z 79 próbek), specyficzność 98% (128 z 131 próbek), dodatnią wartość predykcyjną 96% (67 z 70 badań było poprawnych) oraz ujemną wartość predykcyjną 91% (128 z 140 badań było poprawnych).

	Choroba obecna	Brak choroby	
<b>Nowy test dodatni</b>	A np. 67	B np. 3 Wynik fałszywie dodatni	Jak często wynik dodatni wskazywał chorobę? (70 dodatnich testów; 67 poprawnych)
<b>Nowy test ujemny</b>	C np. 12 Wynik fałszywie ujemny	D np. 128	Jak często wynik ujemny wykluczał chorobę? (140 ujemnych testów; 128 poprawnych)
	Jak często próbka dodatnia była rozpoznana jako dodatnia? (79 dodatnich próbek; 67 z nich jest poprawnie rozpoznanych)	Jak często próbka ujemna była rozpoznana jako ujemna? (131 ujemnych próbek; 128 z nich jest poprawnie rozpoznanych)	



**Ryc. 2a** Przykład krzywej ROC testu, który posiada wysoki stopień czułości i swoistości, a jednocześnie skutecznie łączy je ze sobą. Pole pod krzywą w tym wypadku jest bardzo zbliżone do pola całego obszaru wykresu, oznaczając 1.



**Ryc. 2b** Przykład badania, które uzyskuje jedynie wysoki poziom swoistości kosztem znacznego zmniejszenia czułości. Zaznaczony punkt określa najlepsze połączenie tych dwóch cech. Pole pod krzywą jest znacznie mniejsze niż na rycinie 2a

## Prezentacja wyników

Podstawowa charakterystyka przedstawiona wcześniej opisuje ogólną jakość badania diagnostycznego. Jednakże może być prezentowana na różne sposoby. Oprócz wartości predykcyjnych, istnieje jeszcze tak zwany „współczynnik prawdopodobieństwa” lub „iloraz szans”, który oznacza proporcje prawdopodobieństwa. Czułość i swoistość mogą być przedstawione na tak zwanej krzywej ROC (ang. receiver operating characteristics), w zależności od przyjętych wartości cut-off. W niektórych przypadkach, gdy ma to zastosowanie, obliczany jest również obszar pod krzywą: im bliżej wynik jest równy 1, tym większa jest szansa, że jednocześnie można uzyskać wysokie wartości czułości i swoistości.

### Krzywa ROC

Krzywa ROC (ang. Receiver Operating Characteristic): wykres czułości versus „1 swoistość” umożliwia odczytanie zakresów, w których można zoptymalizować czułość i swoistość danego testu. Przekątna (dwusieczna kąta) osi wskazuje 50% szans występowania choroby. Badania znajdujące się w pobliżu tej linii w ujęciu diagnostycznym nie są dużo bardziej wartościowe niż rzut monetą. Pole pod krzywą (AUC, ang. Area Under the Curve) pod przekątną jest równe 0,5. AUC testów, których krzywe znajdują się w górnym lewym rogu, jest bliskie 1 i pokazują współwystępującą wysoką czułość i swoistość.

Jednakże, w wielu przypadkach (np. w badaniach przesiewowych) nie jest konieczne znalezienie optimum całej krzywej, ponieważ wystarczający może być np. wysoki poziom czułości.

Opisane sposoby prezentacji wyników ostatecznie bazują na tych samych danych, więc mogą się wzajemnie zastępować.

### Źródła

- [1] **Jaltman DG.** (1991): *Practical Statistics for Medical Research*. London: Chapman & Hall.
- [2] **Kirkwood BR.** (1988): *Essentials of medical Statistics*. London: Blackwell Science Ltd.