

Podstawy racjonalnego wyboru testów diagnostycznych w alergologii

The principles of rational selection of diagnostic tests in allergology

ALEKSANDRA GREGORIUS, RADOSŁAW ŚPIEWAK

Uniwersytet Jagielloński Collegium Medicum, Zakład Dermatologii Doświadczalnej i Kosmetologii

Praca wykonana w ramach grantu UJ dla młodych naukowców nr K/DSC/000780 pt. „Optymalizacja testu aktywacji bazofilów w diagnostyce i monitorowaniu immunoterapii alergii na roztocze kurzu domowego”

Streszczenie

Wyniki badań laboratoryjnych oraz testów wykonywanych u pacjenta w znacznym stopniu wpływają na późniejsze etapy diagnostyki i leczenia pacjenta. Dlatego niezwykle ważne jest prawidłowe dobranie testu już na najwcześniejszych etapach diagnostyki oraz uświadomienie sobie, że każdy test – zarówno wykonany w laboratorium, jak i przez samego lekarza, może wiązać się z otrzymaniem wyników fałszywie dodatnich oraz fałszywie ujemnych. W wyborze najlepszego testu niezbędne jest zrozumienie pojęć opisujących skuteczność testu diagnostycznego, takich jak czułość, swoistość, dodatnia i ujemna wartość predykcyjna, wskaźnik wiarygodności dla testu dodatniego i ujemnego, powtarzalność oraz odtwarzalność. W niniejszym artykule przedstawiono definicje tych istotnych pojęć wraz z przykładami z praktyki alergologicznej.

Słowa kluczowe: *alergia, test diagnostyczny, skuteczność, czułość, swoistość, wartości predykcyjne*

Summary

Laboratory and in vivo test results greatly influence further diagnosis and treatment. Therefore, it is essential to select appropriate tests at early stages of the diagnostic process and keep in mind that each test can produce false positive and false negative results. The selection of an optimal test requires a thorough understanding of the terms describing the efficacy of a diagnostic test, such as sensitivity, specificity, positive and negative predictive value, likelihood ratios for a positive and negative test. The present review reports definitions of the relevant terms, along with examples from the allergy practice.

Keywords: *allergy, diagnostic test, efficacy, sensitivity, specificity, predictive value*

© *Alergia Astma Immunologia* 2013, 18 (4): 221-230

www.alergia-astma-immunologia.eu

Przyjęto do druku: 14.03.2013

Adres do korespondencji / Address for correspondence

dr hab. med. Radosław Śpiewak, profesor UJ
ul. Medyczna 9, 30-688 Kraków
tel. 12 620 58 30
e-mail: radoslaw.spiewak@uj.edu.pl

W codziennej praktyce alergolog może wybierać spośród ponad stu ekstraktów alergenowych do testów punktowych i śródskórnych oraz ponad 500 substancji (haptenu) do testów płatkowych, do tego dochodzą dziesiątki testów na swoiste IgE w ofertach laboratoriów analitycznych. Aby wybrać właściwy test oraz prawidłowo go zinterpretować, lekarz powinien być świadomy zalet i wad poszczególnych metod diagnostycznych, umieć wskazać oczekiwane korzyści z wykonania danego testu oraz prawidłowo zinterpretować wyniki w odniesieniu do stanu klinicznego konkretnego pacjenta. Przed zleceniem każdego badania alergolog powinien odpowiedzieć sobie na kilka istotnych pytań [1-3]:

- Jaki jest powód zlecenia testu?
- Jakie mogą być konsekwencje nie zlecenia testu?
- Jak interpretowany jest wynik testu?

- Czy test ma odpowiednią skuteczność w różnicowaniu osób uczulonych od nieuczulonych?
- Czy wynik testu wpłynie na zmianę diagnozy, prognozy bądź terapii?
- Czy wynik zapewni lepsze zrozumienie procesu chorobowego toczącego się u pacjenta?
- Jak wyniki testu wpłyną na życie pacjenta?

W wyborze najlepszego w określonej sytuacji klinicznej testu diagnostycznego pomocne jest dobre zrozumienie podstawowych pojęć opisujących wiarygodność i skuteczność testu (tab. I) [1,2,4-9]. W dobrej praktyce klinicznej niezbędne jest uświadomienie sobie, że każdy test może być źródłem wyników fałszywie dodatnich i fałszywie ujemnych (tab. II). Ryzyko wystąpienia takich błędów opisuje się zazwyczaj za pomocą wskaźników czułości i swoistości testu.

Tabela I. Definicje terminów przydatnych w określaniu skuteczności testu laboratoryjnego

| Termin polski | Termin angielski | Krótką definicja | Wzór |
|--|--|---|---|
| Czułość | <i>Sensitivity</i> | Zdolność testu do wykrywania osób rzeczywiście uczulonych | $\frac{TP}{(TP + FN)} \times 100\%$ |
| Swoistość (specyficzność) | <i>Specificity</i> | Zdolność testu do wykrywania osób rzeczywiście nieuczulonych | $\frac{TN}{(TN + FP)} \times 100\%$ |
| Dodatnia wartość predykcyjna | <i>Positive Predictive Value (PPV)</i> | Procent pacjentów z dodatnim wynikiem testu, którzy rzeczywiście są uczuleni | $\frac{TP}{(TP + FP)} \times 100\%$ |
| Ujemna wartość predykcyjna | <i>Negative predictive value (NPV)</i> | Procent pacjentów z ujemnym wynikiem testu, którzy rzeczywiście nie są uczuleni | $\frac{TN}{(TN + FN)} \times 100\%$ |
| Skuteczność testu diagnostycznego | <i>Efficiency (Accuracy)</i> | Procent pacjentów prawidłowo zakwalifikowanych przez test diagnostyczny jako uczuleni lub nieuczuleni | $\frac{(TP + TN)}{(TP + FP + FN + TN)} \times 100\%$ |
| Wskaźnik wiarygodności testu dla wyniku dodatniego | <i>Likelihood ratio for a positive test (LR+)</i> | Prawdopodobieństwo otrzymania wyniku dodatniego testu u osoby uczulonej w stosunku do prawdopodobieństwa otrzymania wyniku dodatniego testu u osoby nieuczulonej | $\frac{\text{Czułość}}{1 - \text{Swoistość}}$ |
| Wskaźnik wiarygodności testu dla wyniku ujemnego | <i>Likelihood ratio for a negative test (LR-)</i> | Prawdopodobieństwo otrzymania wyniku ujemnego u osoby uczulonej w stosunku do prawdopodobieństwa otrzymania wyniku ujemnego u osoby nieuczulonej | $\frac{1 - \text{Czułość}}{\text{Swoistość}}$ |
| „Złoty standard” | „ <i>Gold standard</i> ” | Pojęcie arbitralne – standard diagnostyczny przyjmowany, jako punkt odniesienia („prawda absolutna”) | Przyjmowany umownie przez badacza |
| Punkt odcięcia | <i>Cut-off value</i> | W przypadku wyników o wartościach ciągłych: punkt, powyżej którego wynik uznajemy za patologiczny | Określany za pomocą krzywych ROC (<i>Receiver-Operator Characteristics</i>) |
| Istotność kliniczna testu | <i>Clinical Relevance</i> | Znaczenie wyniku dla aktualnej choroby (problemu) pacjenta | |
| Powtarzalność* | <i>Reproducibility (Repeatability, Recurrence)</i> | Zgodność niezależnych wyników oznaczenia z tego samego materiału, otrzymanych za pomocą tej samej metody analitycznej, w identycznych warunkach (ten sam analityk, to samo laboratorium, ta sama aparatura, itp.) | |
| Odtwarzalność* | <i>Reproducibility (Interlaboratory agreement)</i> | Zgodność niezależnych wyników oznaczenia z tego samego materiału, otrzymanych za pomocą tej samej metody analitycznej, lecz w różnych laboratoriach, przez różnych analityków, na różnych przyrządach, w różnych okresach | |

TP (wynik prawdziwie dodatni) – wynik testu jest dodatni, u pacjenta występuje uczulenie; TN (wynik prawdziwie ujemny) – wynik testu jest ujemny, u pacjenta nie występuje uczulenie; FP (wynik fałszywie dodatni) – wynik testu dodatni, u pacjenta nie występuje uczulenie; FN (wynik fałszywie ujemny) – wynik testu ujemny, a pacjent jest uczulony

*W literaturze źródłowej terminy „powtarzalność” i „odtwarzalność” są często traktowane jako synonimy

Czułość to zdolność testu do identyfikacji osób rzeczywiście uczulonych. Testy o wysokiej czułości obarczone są wyższym ryzykiem wyników fałszywie dodatnich („wykrywanie” nieistniejących alergii), natomiast wynik ujemny takiego testu z wysokim prawdopodobieństwem przemawia za rzeczywistym brakiem uczulenia u badanego. Test o wysokiej czułości jest szczególnie przydatny w wykluczeniu uczulenia na dany alergen [2,10-12].

Przykład 1: W badaniu z udziałem 71 piekarzy z objawami uczulenia na mąkę pszenną, za „złoty standard” przyjęto wynik testu prowokacji mąką (prowokacja oskrzelowa, donosowa lub ekspozycja w miejscu pracy). U wszystkich pacjentów przeprowadzono również punktowe testy skórne (SPT), których wynik został odniesiony do wyniku testu prowokacji. Uzyskane wyniki przedstawiały się następująco: prawdziwie dodatnie SPT u 25 badanych, prawdziwie ujemne u 25, fałszywie dodatnie u 9, oraz fałszywie ujemne u 12 badanych. Po podstawieniu tych wartości do wzoru $\{25/(25+12)\} \times 100\%$ otrzymujemy wskaźnik czułości dla SPT wynoszący 67,6% [13].

Przykład 2: U 145 pacjentów pediatrycznych przeprowadzono badania przydatności testu Phadiatop Infant (określenie poziomu IgE swoistego wobec najczęstszych alergenów pediatrycznych) w wykrywaniu atopii. Każdy z pacjentów został uprzednio zdiagnozowany i zakwalifikowany do grupy obciążonych lub wolnych od atopii na podstawie wywiadu osobniczego i rodzinnego, badania fizykalnego oraz poziomu swoistego IgE (Pharmacia CAP System) na wybrane alergeny pokarmowe i wziewne. U 53 (36%) dzieci stwierdzono atopię, pozostałe 96 zostało uznane za wolne od atopii. Liczba wyników prawdziwie dodatnich testu Phadiatop wyniosła 49, prawdziwie ujemnych 90, fałszywie dodatnich 4, a fałszywie ujemnych 2. Po podstawieniu tych wartości do wzoru $\{49/(49+2)\} \times 100\%$, otrzymujemy wskaźnik czułości dla testu Phadiatop wynoszący 96,1% [14].

Swoistość to zdolność testu do identyfikacji osób rzeczywiście nieuczulonych. Ponieważ test o wysokiej swoistości z reguły obarczony jest wyższym ryzykiem uzyskania wyników fałszywie ujemnych, dodatni wynik takiego testu z wysokim prawdopodobieństwem przemawia za obecnością uczulenia. Dlatego test o wysokiej swoistości jest szczegó-

nie przydatny w potwierdzaniu uczulenia na dany alergen (np. w orzecznictwie) [2,10,15-17].

Przykład 3: W badaniach określających skuteczność kilku testów diagnostycznych w alergii na lateks wzięto udział 135 pacjentów. Na potrzeby dalszych analiz za punkt odniesienia przyjęto dane z wywiadu (występowanie objawów po kontakcie z lateksem zadeklarowane w kwestionariuszu), na podstawie których nadwrażliwość na lateks stwierdzono u 29 badanych. Dla testu prowokacji rękawiczkami lateksowymi uzyskano wyniki prawdziwie dodatnie u 23 pacjentów, prawdziwie ujemne u 82, fałszywie dodatnie u 6, fałszywie ujemne u 24 badanych. Po podstawieniu tych wartości do wzoru na swoistość $\{82/(82+6)\} \times 100\%$ otrzymujemy wskaźnik swoistości dla testu prowokacji rękawiczkami lateksowymi wynoszący 93,2% (wyliczenia własne na podstawie wyników podanych w pracy [18]).

Przykład 4: W badaniach przeprowadzonych w 2005 roku z udziałem 46 osób uczulonych na lateks oraz 33 osób nieuczulonych, za punkt odniesienia przyjęto połączenie wywiadu, wyników punktowych testów skórnych lub obecności specyficznego IgE. Wszyscy pacjenci zostali następnie przebadani za pomocą cytometrycznego testu aktywacji bazofilów (BAT). Liczba wyników prawdziwie dodatnich wyniosła 39, prawdziwie ujemnych 29, fałszywie dodatnich 4, a fałszywie ujemnych 7. Po podstawieniu wartości do wzoru na swoistość $\{29/(29+4)\} \times 100\%$ otrzymujemy wskaźnik swoistości BAT wynoszący 87,9% [19].

Czułość i swoistość testu można oszacować na podstawie stosunkowo prostych badań klinicznych. „Idealny” test o 100% czułości i 100% swoistości kwalifikowałby prawidłowo wszystkich pacjentów, jednak zgodnie z teorią Bayesa żaden pojedynczy test nie osiąga takiej skuteczności. Dodatkowo Krieg zwrócił uwagę na notoryczny brak w publikacjach i ulotkach załączonych do odczynników szczegółowych informacji na temat populacji, w jakiej zebrano dane wykorzystane do wyliczenia swoistości i czułości. W przypadku zaawansowanej choroby test może prawidłowo kwalifikować pacjentów z blisko 100% czułością, jednak u pacjentów w początkowych stadiach choroby może on wykazywać zaledwie 20% czułość, co będzie skutkowało nieprawidłowym rozpoznaniem aż u 80% chorych [1]. Błędem jest zatem interpretacja wyników testu na podstawie

Tabela II. Możliwe wyniki testu diagnostycznego

| | Uczuleni | Nieuczuleni | |
|--------------|---|---|--|
| | Według wyniku testu uznanego za „złoty standard” | | |
| Test dodatni | Prawdziwie dodatnie (<i>true positive</i> , TP) | Fałszywie dodatnie (<i>false positive</i> , FP) | Całkowita liczba dodatnich testów (TP+FP) |
| Test ujemny | Fałszywie ujemne (<i>false negative</i> , FN) | Prawdziwie ujemne (<i>true negative</i> , TN) | Całkowita liczba ujemnych testów (FN+TN) |
| | Całkowita liczba uczulonych (TP+FN) | Całkowita liczba nieuczulonych (FP+TN) | Całkowita liczba badanych (TP+FP+FN+TN) |

TP (wynik prawdziwie dodatni) – wynik testu jest dodatni, u pacjenta występuje uczulenie; TN (wynik prawdziwie ujemny) – wynik testu jest ujemny, u pacjenta nie występuje uczulenie; FP (wynik fałszywie dodatni) – wynik testu dodatni, u pacjenta nie występuje uczulenie; FN (wynik fałszywie ujemny) – wynik testu ujemny, a pacjent jest uczulony

samej czułości bądź swoistości przy braku informacji o kryteriach włączenia do grupy badanej. Dodatkowo opieranie się na czułości i swoistości testu skutkuje pominięciem istotnego czynnika, jakim jest częstość występowania badanej cechy w określonej populacji. Można to zilustrować przykładem hipotetycznej grupy 1000 osób, w której częstość uczulenia na alergen X wynosi 40%, a na alergen Y - 2%. Różnice skuteczności dwóch testów o identycznych wskaźnikach czułości i swoistości wynoszących w każdym przypadku po 95% ilustruje tabela III, która uzmysławia, że w przypadku częstego (40%) uczulenia na alergen X odsetek osób prawdziwie uczulonych wśród osób z dodatnim wynikiem testu wyniesie 92,7%, natomiast w przypadku rzadkiego (2%) uczulenia na alergen Y odsetek ten wyniesie zaledwie 27,9%. Sytuacja, w której 2 spośród 3 osób z dodatnim wynikiem testu w istocie nie są uczulone wydaje się częsta w codziennej praktyce, jednak trudna do zaakceptowania z punktu widzenia lekarza klinicysty. Dlatego w praktyce lekarskiej pożądana jest ocena testu na podstawie wartości predykcyjnych, które uwzględniają częstość występowania badanej cechy w wybranej populacji. Dodatnia i ujemna wartość predykcyjna opisują prawdopodobieństwo faktycznej obecności lub nieobecności uczulenia u pojedynczego pacjenta z danym wynikiem testu diagnostycznego. Powracając do przykładu z tabeli III, wartości predykcyjne dodatnie i ujemne wyniosłyby 92,7% i 96,6% przy częstości uczulenia 40% oraz odpowiednio 27,9% i 99,9% przy częstości 2%. Mimo bez wątpienia większej wartości klinicznej, oszacowanie wartości predykcyjnych wiąże się z koniecznością przeprowadzenia pracochłonnych i kosztownych przekrojowych badań populacyjnych osobno dla każdego testu, alergenu oraz populacji docelowej (rozpoznanie, grupa wiekowa, ekspozycja).

Wartość predykcyjna dodatnia (PPV) określa procent pacjentów, którzy rzeczywiście są uczuleni wśród wszystkich osób z dodatnim wynikiem testu. **Wartość predykcyjna ujemna (NPV)** określa procent pacjentów, którzy faktycznie nie są uczuleni wśród wszystkich osób z ujemnym wynikiem testu [1,2,10,20,21].

Przykład 5: Na bazie międzynarodowych badań nad występowaniem astmy i alergii w dzieciństwie (*International Study of Asthma and Allergies in Childhood, ISAAC*) w trzech hiszpańskich miastach przeprowadzono randomi-

zowane analizy na grupie dzieci w wieku szkolnym (9-12 lat). Do rodziców rozesłano kwestionariusze, a na podstawie odpowiedzi oceniono obecność atopii u dziecka. U 621 dzieci wykonano punktowe testy skórne (SPT) z 8 najczęściej uczulającymi alergenami (*D. pteronyssinus*, *D. farine*, kot, *Alternaria*, mieszanka alergenów pyłków drzew, mieszanka alergenów pyłków traw, pyłek drzewa oliwnego, *Parietaria*). Za marker atopii przyjęto obecność reakcji dodatniej na minimum jeden alergen. Na podstawie odpowiedzi w kwestionariuszu, dzieci zakwalifikowano ponadto do grup z astmą (79), zapaleniem śluzówki nosa i spojówek (102) oraz bez objawów ze strony układu oddechowego (474). U wszystkich dzieci wykonano następnie test Phadiatop™ i wyliczono wskaźniki skuteczności dla całej grupy oraz dla grup z poszczególnymi objawami. Dodatnia oraz ujemna wartość predykcyjna tego testu wyniosły odpowiednio dla całej populacji 72,7% i 92,7%, dla dzieci z objawami astmy 87,5% i 95,6%, dla dzieci z objawami zapalenia śluzówki nosa i spojówek 84,0% i 93,9%, a dla dzieci bez objawów ze strony układu oddechowego odpowiednio 64,3% i 92,4%. Dodatkowo zarówno test Phadiatop™, jak i punktowe testy skórne nieprawidłowo sugerowały atopię u 10% dzieci faktycznie zdrowych (wyniki fałszywie dodatnie) [22,23].

Przykład 6: Do badań nad skutecznością testu Phadiatop™ w populacji dorosłych wylosowano 469 pacjentów (18-92 lata, mediana 54). Do badanych rozesłano kwestionariusze w celu zebrania danych demograficznych oraz informacji o narażeniach i obecności objawów ze strony górnych dróg oddechowych. Dodatkowo u wszystkich wykonano skórne testy punktowe z alergenami roztoczy kurzu domowego, pyłkami traw i drzew, pleśni, sierścią kota i psa. Test Phadiatop™ został wykonany u 465 spośród 469 osób, u 4 wykonanie testu nie było możliwe ze względów technicznych. Przyjmując za punkt odniesienia wynik punktowych testów skórnych, w teście Phadiatop™ uzyskano 85 wyników prawdziwie dodatnich, 32 fałszywie dodatnie, 313 prawdziwie ujemnych oraz 35 fałszywie ujemnych. Dla pacjentów deklarujących w kwestionariuszu objawy ze strony układu oddechowego (n=221) wyniki przedstawiały się następująco: prawdziwie dodatnie 61, fałszywie dodatnie 12, prawdziwie ujemne 132 oraz fałszywie ujemne u 16 pacjentów. Na podstawie tych wyników obliczono wartości

Tabela III. Hipotetyczne wyniki dla testu o czułości i swoistości 95% w zależności od częstości uczulenia na dany alergen w populacji złożonej z 1000 osób

| Wynik | Uczulenie na alergen X (40% populacji) | Uczulenie na alergen Y (2% populacji) |
|--------------------------|---|---------------------------------------|
| Uczuleni | 400 | 20 |
| Nieuczuleni | 600 | 980 |
| Dodatnie wyniki testu | 410 | 68 |
| Ujemne wyniki testu | 590 | 932 |
| Prawdziwie dodatnie (TP) | 380 (38%) | 19 (1,9%) |
| Fałszywie dodatnie (FP) | 30 (3%) | 49 (4,9%) |
| Prawdziwie ujemne (TN) | 570 (57%) | 931 (93,1%) |
| Fałszywie ujemne (FN) | 20 (2%) | 1 (0,1%) |

predykcyjne dodatnie i ujemne, które wyniosły odpowiednio 72,6% i 89,9% dla całej badanej populacji dorosłych oraz 83,5% i 89,1% dla dorosłych z objawami ze strony układu oddechowego [24,25].

Skuteczność testu diagnostycznego to procent pacjentów prawidłowo zakwalifikowanych przez test diagnostyczny do grupy uczulonych bądź nieuczulonych [26].

Przykład 7: W badaniach nad skutecznością testu diagnostycznego do samodzielnego wykonania wykrywającego obecność alergii na nikiel (Nixema®, Mekos Laboratories) wzięło udział 191 badanych. Za „złoty standard” przyjęto wynik testów płatkowych jednocześnie wykonywanych w klinice dermatologicznej. Po odczycie samodzielnie wykonanego testu przez pacjentów i lekarzy oraz testu płatkowego przez lekarzy stwierdzono 33 wyniki prawdziwie dodatnie, 132 prawdziwie ujemne, 13 fałszywie dodatnich oraz 13 fałszywie ujemnych. Po podstawieniu wartości do wzoru na skuteczność testu diagnostycznego $\{33+132/(33+132+13+13)\} \times 100\%$, otrzymujemy wskaźnik skuteczności dla testu Nixema® wynoszący 86,4% (wyliczenia własne na podstawie wyników podanych w pracy [27]).

Przykład 8: W celu określenia skuteczności testu RAST w wykrywaniu alergii na tymotkę przeprowadzono badania z udziałem 69 pacjentów. Na podstawie przyjętego przez autorów „złotego standardu”, jakim była prowokacja donosowa lub dospojówkowa, alergię na tymotkę stwierdzono u 36 badanych. W teście RAST uzyskano 34 wyniki prawdziwie dodatnie, 23 prawdziwie ujemne, 10 fałszywie dodatnich oraz 2 fałszywie ujemne. Po podstawieniu tych wartości do wzoru na skuteczność testu diagnostycznego $\{34+23/(34+23+10+2)\} \times 100\%$, otrzymujemy dla testu RAST wskaźnik skuteczności wynoszący 82,6% (wyliczenia własne na podstawie wyników podanych w pracy [28]).

Wskaźnik wiarygodności dla testu dodatniego (LR+) to prawdopodobieństwo otrzymania wyniku dodatniego u osoby uczulonej w stosunku do prawdopodobieństwa otrzymania wyniku dodatniego testu u osoby nieuczulonej. Wskaźnik LR+ może przyjmować wartości od 0 do nieskończoności, wyższe wartości przemawiają na korzyść wiarygodności testu dodatniego.

Wskaźnik wiarygodności dla testu ujemnego (LR-) to prawdopodobieństwo otrzymania wyniku ujemnego u osoby uczulonej w stosunku do prawdopodobieństwa otrzymania wyniku ujemnego u osoby nieuczulonej. Wartości LR- zbliżone do 0 sugerują wyższą wiarygodność testu ujemnego. Wskaźniki wiarygodności są rzadko stosowane ze względu na mało intuicyjną i trudną interpretację.

Przykład 9: W badaniach przeprowadzonych w celu określenia czułości i swoistości testów śródskórnych u 37 dorosłych z objawami sezonowego alergicznego nieżytku nosa w wywiadzie, za „złoty standard” autorzy przyjęły wynik prowokacji donosowej z alergenem pyłku tymotki. Wyliczone wskaźniki czułości i swoistości dla testów śródskórnych wyniosły odpowiednio 93,3% (0,933) oraz 86,4% (0,864). Po podstawieniu tych wartości do wzoru na wskaźnik wiarygodności testu dodatniego $\{0,933/(1-0,864)\}$ otrzymujemy wskaźnik LR+ wynoszący 6,86. Po podstawieniu do wzoru na wskaźnik wiarygodności testu ujem-

nego $\{(1-0,933)/0,864\}$ otrzymujemy wskaźnik LR- równy 0,08 (wyliczenia własne na podstawie wyników podanych w pracy [29]).

Przykład 10: W badaniach określających skuteczność testu Allergodip® (Allergopharma) w wykrywaniu IgE swoistych wobec lateksu wzięło udział 135 pacjentów. Za punkt odniesienia przyjęto wynik testu ImmunoCAP®, za pomocą którego u 69 badanych potwierdzono alergię na lateks. Na podstawie tych wyników wyliczono wskaźnik czułości i swoistości dla testu Allergodip®, który wyniósł odpowiednio 85,1% (0,851) oraz 75,0% (0,750). Po podstawieniu tych wartości do wzoru na wskaźnik wiarygodności testu dodatniego $\{0,851/(1-0,750)\}$ otrzymujemy wskaźnik LR+ wynoszący 3,4. Po podstawieniu do wzoru na wskaźnik wiarygodności testu ujemnego $\{(1-0,851)/0,750\}$ otrzymujemy wskaźnik LR- równy 0,20 (wyliczenia własne na podstawie wyników podanych w pracy [18]).

Zgodność testów opisuje się zwykle za pomocą dwóch terminów: powtarzalności i odtwarzalności. Powtarzalność to zgodność wyników niezależnych badań tego samego materiału przeprowadzonych przez tego samego badacza w identycznych warunkach za pomocą tej samej metody. W diagnostyce laboratoryjnej można ponadto spotkać się z terminem odtwarzalności, który odnosi się do możliwych rozbieżności wyników badania tego samego materiału w różnych laboratoriach. W publikacjach oba terminy nierzadko są traktowane jako synonimy [30-36].

Przykład 11: W 2008 roku Ronchetti i wsp. przeprowadzili badania z udziałem 227 dzieci w celu ustalenia powtarzalności wyników atopowych testów płatkowych (APT) z komercyjnymi i świeżymi alergenami pokarmowymi (mleko krowie, jaja kurze, pomidor, mąka pszenna) oraz komercyjnymi alergenami wziewnymi (*Dermatophagoides pteronyssinus*, mieszanka traw). Testy zostały wykonane w dwóch powtórzeniach na lewej i prawej stronie pleców tych samych pacjentów. Uzyskane wyniki sugerują niską powtarzalność atopowych testów płatkowych w przypadku alergenów pokarmowych zarówno komercyjnych, jak i świeżych. Odnotowana przez autorów powtarzalność APT z natywnymi pokarmami mieściła się w zakresie od 38% (pomidor) do maksymalnie 81% (mleko). Natomiast w przypadku aeroalergenów autorzy stwierdzili 100% powtarzalność [37].

Przykład 12: W celu porównania odtwarzalności wyników testów płatkowych aplikowanych przy użyciu technik zaproponowanych przez dwóch producentów: testu T.R.U.E. (Mekos) oraz aplikacji haptentów za pomocą komór IQ (Chemotechnique), u 207 pacjentów przeprowadzono analizę wyników równoczesnego badania oboma systemami z 21 substancjami testowymi (hapteny lub mieszanki haptentów). Całkowita zgodność pomiędzy obiema metodami wyniosła 64,4%. W odniesieniu do testu z kobaltem autorzy stwierdzili pominięcie istotnych klinicznie odczynów u 78% pacjentów w przypadku zastosowania T.R.U.E. testu oraz 4% pacjentów w przypadku zastosowania komór IQ. Jednocześnie T.R.U.E. test wykrył wszystkie istotne klinicznie odczyny na Quaternium 15, w odróżnieniu od aplikacji haptenu w komorach IQ, za pomocą których nie wykryto 50% reakcji. Korelacja dodatnich wyników znacząco różniła się dla poszczególnych haptentów i była najwyższa (100%)

dla kalafonii, mieszanki parabenów oraz mieszanki merkaptanów. Brak korelacji w wynikach dodatnich odnotowano dla klio chinolu i lanoliny (alkohole wełny). W przypadku użycia komór IQ uzyskano większą liczbę wyników fałszywie dodatnich, w przypadku testu T.R.U.E. większą liczbę wyników fałszywie ujemnych. Z punktu widzenia praktyka, uzyskanie wyników fałszywie ujemnych przy użyciu testu T.R.U.E. wiąże się z większym ryzykiem pominięcia istotnych klinicznie wyników dla haptenów takich jak chrom, substancje zapachowe, kobalt i balsam peruwiański, które zostałyby wykryte w przypadku zastosowania komór IQ [38].

Przykład 13: W celu porównania wyników pomiaru swoistego IgE trzema różnymi metodami laboratoryjnymi, surowicę od 50 pacjentów pediatrycznych przebadano w trzech różnych laboratoriach diagnostycznych z sześcioma wybranymi alergenami (jajko, mleko, orzech ziemny, sierść kota, brzoza oraz *Dermatophagoides farinae*) [39]. Wyniki uzyskane przy równoległym zastosowaniu metod Turbo-MP (Agilent Technologies Co) oraz Immulite 2000 (Siemens Medical Solutions Diagnostics) zostały odniesione do wyników uzyskanych z zastosowaniem metody ImmunoCAP (Phadia). Dla metody Turbo-MP badacze stwierdzili zgodność wyników tylko dla alergenu jaja, nieznaczne różnice dla alergenów mleka i orzecha ziemnego, natomiast znaczne zaniżenie wyników stwierdzono w przypadku brzozy i *D. farinae*. W metodzie Immulite wyniki pomiaru IgE swoistego dla wszystkich alergenów były zawyżone w porównaniu do wyników uzyskanych w metodzie ImmunoCAP. Autorzy zwrócili uwagę na możliwe różnice w budowie alergenów dystrybuowanych przez producentów poszczególnych metod pomiarowych. Dopiero po konsultacjach z firmami ustalono, że pod nazwą „alergen brzozy” kryły się ekstrakty pyłków pochodzące z różnych gatunków brzozy, jednak nie było o tym wzmianki w informacjach dołączanych do produktów.

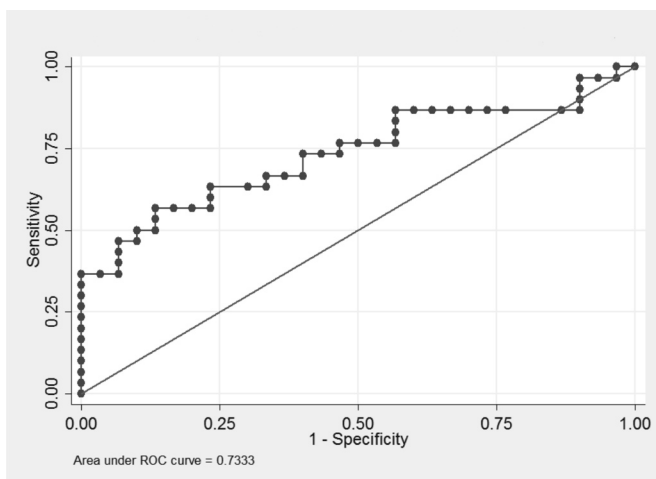
Krzywe ROC (Receiver-Operator Characteristics)

W przypadku, gdy wyniki testów diagnostycznych mają charakter zmiennych ciągłych (np. poziom swoistych IgE lub średnica bąbla w teście punktowym) i mogą przybrać

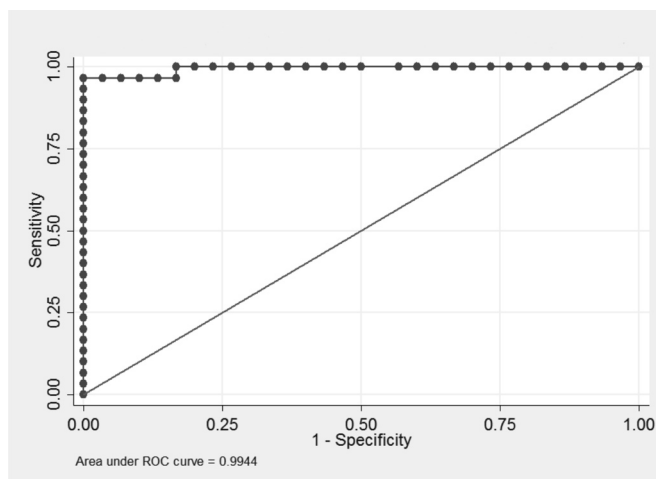
dowolną wartość w zakresie pomiarowym testu, lekarz staje przed dylematem, od jakiej wartości wynik może być uznany za dodatni. Narzędziem wyboru optymalnych punktów odcięcia wyników pozytywnych od negatywnych, a zatem optymalizacji czułości i swoistości testu jest wypracowanie krzywych ROC (Receiver-Operator Characteristics Curves). Najczęściej stosuje się krzywą ROC przedstawiającą zależność częstości występowania wartości prawdziwie dodatnich (czułość) od wartości fałszywie dodatnich (1-swoistość) [40-42]. Analiza krzywej ROC pozwala na ustalenie najbardziej akceptowalnego kompromisu między czułością a swoistością testu, a tym samym osiągnięcie największej skuteczności w wykrywaniu uczulenia na konkretny alergen lub haptenu. Zdolność testu do prawidłowego różnicowania osób uczulonych od nieuczulonych rośnie w miarę zbliżania się krzywej do lewego górnego rogu wykresu. Rycina 1 przedstawia krzywą ROC dla testu o niskiej skuteczności (krzywa oddalona od lewego górnego rogu wykresu), natomiast rycina 2 przedstawia krzywą ROC testu o wysokiej skuteczności (krzywa przylega do rogu).

Punkt odcięcia testu nie jest stałą cechą przypisaną do danej metody diagnostycznej, lecz zależy od szeregu czynników, takich jak rodzaj alergenu lub haptenu (naturalny czy rekombinowany, mieszanka alergenów czy pojedynczy epitop), populacja docelowa (wiek badanych, choroba główna i stany współistniejące) oraz przyjęty punkt odniesienia, czyli tzw. „złoty standard” [43-49].

Przykład 14: Plebani i wsp. w 1996 roku przeprowadzili szczegółowe analizy punktów odcięcia dla poziomu IgE (Immuno CAP) swoistego wobec 5 alergenów wziewnych u pacjentów z alergią oddechową. Przyjmując za punkt odniesienia kompilację historii choroby, danych klinicznych i wyników SPT autorzy stwierdzili, że optymalny punkt odcięcia powinien być ustalany dla każdego alergenu indywidualnie. Przykładowe wyniki oraz zależność czułości i swoistości od przyjętego punktu odcięcia ilustruje tabela IV. Jak widać czułość i swoistość są cechami odwrotnie proporcjonalnymi, co oznacza, że zmieniając parametry testu tak, by zwiększyć jeden wskaźnik, musimy liczyć się z obniżeniem drugiego [50].



Ryc. 1. Przykład krzywej ROC wskazującej na niską skuteczność testu diagnostycznego (dane własne)



Ryc. 2. Przykład krzywej ROC wskazującej na wysoką skuteczność testu diagnostycznego (dane własne)

„Złoty standard” a skuteczność testu

W ocenie skuteczności testu niezmiernie ważne jest dobranie odpowiedniego punktu odniesienia, do którego będziemy porównywać wyniki ocenianej metody. W alergologii rzadko mamy do czynienia z pojedynczym testem, który jednoznacznie kwalifikowałby pacjentów jako uczulonych lub nieuczulonych. Najbliższe takiemu ideałowi wydają się prowokacje alergenowe, co czyni z nich preferowany „złoty standard” przy ocenie innych metod diagnostycznych. Nie zawsze jednak prowokacja jest wykonalna lub akceptowalna ze względów etycznych. W ostatnich latach zwrócono ponadto uwagę na problem lokalnego alergicznego nieżytku nosa, charakteryzującego się miejscową produkcją swoistych IgE przy ich nieobecności w krwi krążącej, co się objawia dodatnim wynikiem w teście prowokacji donosowej przy ujemnych wynikach oznaczeń IgE swoistych w surowicy i punktowych testach skórnych [51-53]. Dlatego niekiedy wybiera się inne „złote standardy”, w tym uprzednio poddane walidacji metody diagnostyczne lub ich kombinacje. Należy przy tym pamiętać, że żaden z wybranych „złotych standardów” nie stanowi prawdy absolutnej i wyniki oceny skuteczności testu zależą od wybranej metody referencyjnej [54]. Najbardziej racjonalnym podejściem jest

krytyczne spojrzenie na wyniki badań dodatkowych oraz indywidualne podejście do diagnozy stawianej u każdego pacjenta, której podstawą niezmiennie pozostaje dobrze zebrany wywiad lekarski.

Przykład 15: W celu porównania skuteczności trzech dostępnych testów diagnostycznych do oznaczania IgE swoistego wobec lateksu w surowicy (ImmunoCAP®, AlaSTAT®, HY-TEC®), ich czułość i swoistość oraz wartości predykcyjne oceniono na podstawie dwóch różnych „złotych standardów”: historii choroby oraz wyników punktowych testów skórnych. Uzyskane wyniki pokazują znaczące różnice w ocenie skuteczności poszczególnych testów w zależności od wybranego „złotego standardu” (tab. V) [55].

Przykład 16: W roku 1986 w Szwecji zbadano różnice pomiędzy skutecznością różnych technik diagnostycznych (wywiad, punktowe testy skórne, RAST) oraz ich kombinacji w wykrywaniu alergii na brzozę u 69 pacjentów z nieżytem nosa i/lub spojówek. Za punkt odniesienia przyjęto wynik prowokacji spojówkowej i nosowej. Autorzy doszli do wniosku, że w praktyce klinicznej najlepiej oprzeć się na wyniku testów punktowych oraz poziomie swoistego IgE (tab. VI) [28].

Tabela IV. Zależność czułości i swoistości od przyjętego progu odcięcia dla poszczególnych alergenów (wg [50], zmodyfikowane)

| Alergen | ROC: Punkt odcięcia | Czułość | Swoistość |
|---------------------------------------|---------------------|---------|-----------|
| <i>Dermatophagoides pteronyssinus</i> | 0,40 kUa/l | 96% | 94% |
| | 1,20 kUa/l | 80% | 100% |
| <i>Lolium perenne</i> | 0,36 kUa/l | 100% | 97% |
| | 0,55 kUa/l | 97% | 100% |
| <i>Parietaria officinalis</i> | 0,65 kUa/l | 92% | 98% |
| | 0,35 kUa/l | 100% | 87% |

Tabela V. Czności i swoistości przykładowych testów w zależności od przyjętego „złotego standardu” klinicznego (wg [55], zmodyfikowane)

| Metoda | Analizowany parametr | „Złoty standard”: Wywiad | „Złoty standard”: Wynik skórnych testów punktowych |
|------------|----------------------|-----------------------------|---|
| ImmunoCAP® | Czułość | 75,2% | 76,3% |
| | Swoistość | 90,8% | 96,7% |
| AlaSTAT® | Czułość | 78,6% | 97,2% |
| | Swoistość | 95,4% | 95,0% |
| HY-TEC® | Czułość | 89,8% | 73,3% |
| | Swoistość | 67,5% | 71,4% |

Tabela VI. Zależność czności i swoistości od kombinacji użytych testów diagnostycznych (wg [28], zmodyfikowane)

| Test diagnostyczny | Czułość | Swoistość |
|---------------------|---------|-----------|
| SPT + RAST | 100% | 100% |
| Wywiad + SPT + RAST | 100% | 100% |
| Wywiad + SPT | 97% | 96% |
| SPT | 97% | 94% |
| Wywiad + RAST | 87% | 100% |
| RAST | 82% | 100% |
| Wywiad | 79% | 90% |

Jakość alergenu a skuteczność testu

W punktowych testach skórnych, atopowych testach płatkowych oraz testach *in vitro* wykrywających swoiste IgE wykorzystuje się naturalne ekstrakty alergenów lub rzadziej antygeny rekombinowane. Różnice w zawartości poszczególnych alergenów w wyrobach różnych producentów, brak standaryzacji testów oraz powszechnie uznanych rekomendacji dotyczących rodzaju stosowanych alergenów mogą skutkować rozbieżnością wyników w przypadku wykonania testów u tego samego pacjenta w różnych laboratoriach diagnostycznych lub przez różnych lekarzy. W badaniach przeprowadzonych w Danii wykazano, że spośród dziewięciu komercyjnie dostępnych ekstraktów alergenowych orzecha laskowego do testów skórnych lub oznaczania swoistego IgE, trzy nie zawierały epitopu odpowiedzialnego za występowanie reakcji anafilaktycznych. Na skutek tego u sześciu spośród trzydziestu przebadanych pacjentów (20%) nie wykryto klinicznie istotnej nadwrażliwości na ten alergen [56].

Istotność kliniczna dodatniego testu

Objawy obecne u pacjentów zgłaszających się do poradni alergologicznych rzadko są charakterystyczne na tyle, by lekarz był w stanie zidentyfikować odpowiedzialny alergen lub hapten bez dodatkowych badań. Niggemann i wsp. wykazali, że znajomość poziomów swoistych IgE w surowicy pacjentów zmniejsza odsetek wątpliwych rozpoznań z 86% do 22%, co znacząco poprawia skuteczność diagnostyki już w trakcie pierwszej wizyty pacjenta oraz skraca czas oczekiwania na ostateczne rozpoznanie [57]. Podobnie Rajagopalan i wsp. zaobserwowali, że wykonanie testów płatkowych zwiększa szansę ostatecznego rozpoznania i ponad 20-krotnie skraca czas oczekiwania na ostateczne rozpoznanie [58,59]. Z drugiej strony dodatni wynik testu diagnostycznego nie implikuje, że dany alergen lub hapten jest faktyczną przyczyną bieżących dolegliwości pacjenta

m.in. w związku z możliwością alergii krzyżowej lub odczynem na alergen istotny w przeszłości, z którym aktualnie pacjent nie ma już styczności. Dlatego w przypadku każdego dodatniego wyniku lekarz powinien poszukiwać odpowiedzi na pytanie czy wykryte uczulenie faktycznie jest przyczyną obecnej choroby [60]. Na skuteczność testów wpływa ponadto dobór i liczba testowanych alergenów lub haptentów [61,62].

Podsumowanie

W procesie diagnostycznym rolą laboratorium jest rzetelne wykonanie zleconych badań, przy czym stosowane testy powinny być „dopasowane” do populacji, która zgłasza się do danego lekarza. To samo odnosi się do testów diagnostycznych *in vivo* wykonywanych przez lekarza. Laboratoria stosują przeważnie zestawy diagnostyczne zagranicznych producentów, a normy podane w ulotkach produktów zwykle odnoszą się do populacji w krajach pochodzenia, gdzie między innymi częstość występowania uczulenia na dany alergen może zasadniczo różnić się od populacji polskiej. Również wiek i choroba podstawowa mogą znacząco wpływać na dobór punktów odcięcia dla danego testu diagnostycznego. W odniesieniu do testów *in vivo*, kwestia wiarygodności testów jest tematem zaledwie pojedynczych prac badawczych. Ponieważ wynik testu może rzutować nawet na 70% podejmowanej decyzji medycznej, zrozumienie znaczenia wskaźników określających skuteczność każdej procedury diagnostycznej – zarówno tej stosowanej w laboratorium, jak i testu wykonywanego przez samego lekarza – jest fundamentem prawidłowej oceny stanu pacjenta [2]. W szczególności potrzebne jest opracowanie norm dla populacji, z którą alergolog ma do czynienia w swojej praktyce. Pozwoli to zminimalizować ryzyko zaniechania niezbędnej terapii u osób potrzebujących lub narażenia na niepotrzebne leczenie błędnie zdiagnozowanych pacjentów.

Piśmiennictwo

- Krieg AF. Why are clinical laboratory tests performed. *JAMA* 1975; 233: 76-8.
- Wians FH. Clinical Laboratory Tests: Which, Why, and What do the results Mean? *Lab Medicine* 2009; 40: 105-13.
- Price CP. Evidence-based laboratory medicine: supporting decision-making. *Clin Chem* 2000; 46: 1041-50.
- Śpiewak R. Alergia kontaktowa – diagnostyka i postępowanie. *Alergia Astma Immunologia* 2007; 12: 109-27.
- Akobeng AK. Understanding diagnostic tests 1: Sensitivity, specificity and predictive values. *Acta Paediatrica* 2006; 96: 338-41.
- Diepgen TL, Coenraads PJ. Sensitivity, specificity and positive predictive value of patch testing: the more you test, the more you get? *ESCD Working Party on Epidemiology. Contact Dermatitis* 2000; 42: 315-7.
- Horn PS, Pesce AJ. Reference intervals: an update. *Clin Chim Acta* 2003; 334: 5-23.
- Sunderman FW Jr. Current concepts of “normal values,” “reference values,” and “discrimination values,” in clinical chemistry. *Clin Chem* 1975; 21: 1873-7.
- Wieland E. The IFCC recommendations for determining reference intervals: strengths and limitations. *J Lab Med* 2009; 33: 45-51.
- Galen RS. Letter: Predictive value of laboratory tests. *Am J Cardiol* 1975; 36: 536-8.
- Czarnobilska E, Gregorius A, Porębski G i wsp. Korzyści z wykonywania testu aktywacji bazofilów w kwalifikacji do immunoterapii swoistej w alergii wziewnej. *Przegląd Lekarski* 2012; 69: 1249-53.
- Małolepszy J, Mędrala W, Kuna P i wsp. Standardy w alergologii. Część II. Oznaczanie całkowitego stężenia IgE i alergenowo-swoistych IgE w surowicy. *Przeg Alergol* 2004; 1: 51-9.
- van Kampen V, Rabstein S, Sander I i wsp. Prediction of challenge test results by flour-specific IgE and skin prick test in symptomatic bakers. *Allergy* 2008; 63: 897-902.
- Lau S, Nilsson M, Sulser C i wsp. Use of Phadiatop Infant in diagnosis of specific sensitization in young children with wheeze or eczema. *Pediatr Allergy Immunol* 2008; 19: 337-41.
- Canani RB, Ruotolo S, Auricchio L i wsp. Diagnostic accuracy of the atopy patch test in children with food allergy-related gastrointestinal symptoms. *Allergy* 2007; 62: 738-43.

16. Paganelli R, Ansotegui IJ, Sastre J i wsp. Specific IgE antibodies in the diagnosis of atopic disease. Clinical evaluation of a new in vitro test system, UniCAP, in six European allergy clinics. *Allergy* 1998; 53: 763-8.
17. Ebo DG, Lechkar B, Schuerwegh AJ i wsp. Validation of a two-color flow cytometric assay detecting in vitro basophil activation for the diagnosis of IgE-mediated natural rubber latex allergy. *Allergy* 2002; 57: 706-12.
18. Niggemann B, Wahn U. A new dipstick test (Allergodip®) for in vitro diagnosis of latex allergy - validation in patients with spina bifida. *Pediatr Allergy Immunol* 2000; 11: 56-9.
19. Hemery ML, Arnoux B, Dhivert-Donnadieu H i wsp. Confirmation of the diagnosis of natural rubber latex allergy by the Baso-test method. *Int Arch Allergy Immunol* 2005; 136: 53-7.
20. Nilsson C, Lilja G, Nordlund M i wsp. Phadiatop Infant(®) detects IgE-mediated diseases among pre-school children: a prospective study. *Pediatr Allergy Immunol* 2012; 23: 159-65.
21. Peters RL, Gurrin LC, Allen KJ. The predictive value of skin prick testing for challenge-proven food allergy: a systematic review. *Pediatr Allergy Immunol* 2012; 23: 347-52.
22. Garcia-Marcos L, Castro-Rodriguez JA, Suarez-Varela MM i wsp. A different pattern of risk factors for atopic and non-atopic wheezing in 9-12-year-old children. *Pediatr Allergy Immunol* 2005; 16: 471-7.
23. Garcia-Marcos L, Sanchez-Solis M, Martinez-Torres AE i wsp. Phadiatop™ compared to skin-prick test as a tool for diagnosing atopy in epidemiological studies in schoolchildren. *Pediatr Allergy Immunol* 2007; 18: 240-4.
24. Vidal C, Gude F, Boquete O i wsp. Evaluation of the phadiatop test in the diagnosis of allergic sensitization in a general adult population. *J Investig Allergol Clin Immunol* 2005; 15: 124-30.
25. Vidal C, Boquete O, Gude F i wsp. High prevalence of sensitization to storage mites in general adult population. *Allergy* 2004; 59: 401-5.
26. Choi IS, Koh YI, Koh JS, Lee MG. Sensitivity of the skin prick test and specificity of the serum-specific IgE test for airway responsiveness to house dust mites in asthma. *J Asthma* 2005; 42: 197-202.
27. Josefson A, Svensson A, Färm G. Validation of self-testing as a method to estimate the prevalence of nickel allergy. *Acta Derm Venereol* 2011; 91: 526-30.
28. Petersson G, Dreborg S, Ingestad R. Clinical history, skin prick test and RAST in the diagnosis of birch and timothy pollinosis. *Allergy* 1986; 41: 398-407.
29. Krouse JH, Sadrazodi K, Kerswill K. Sensitivity and specificity of prick and intradermal testing in predicting response to nasal provocation with timothy grass antigen. *Otolaryngol Head Neck Surg* 2004; 131: 215-9.
30. Jiang XD, Li GY, Dong Z, Zhu DD. Correlation analysis of two serum-specific immunoglobulin E test systems and skin-prick test in allergic rhinitis patients from northeast China. *Am J Rhinol Allergy* 2011; 25: 116-9.
31. Bruze M, Isaksson M, Edman B i wsp. A study on expert reading of patch test reactions: inter-individual accordance. *Contact Dermatitis* 1995; 32: 331-7.
32. Bodtger U, Jacobsen CR, Poulsen LK, Malling HJ. Long-term repeatability of the skin prick test is high when supported by history or allergen-sensitivity tests: a prospective clinical study. *Allergy* 2003; 58: 1180-6.
33. Brasch J, Henseler T, Aberer W i wsp. Patch testing of nickel sulfate and potassium dichromate with a standardized ready-to-use test system gives highly reproducible results: a double-blind multicentre study. *Acta Derm Venereol* 2001; 81: 122-4.
34. Bourke JF, Batta K, Prais L i wsp. The reproducibility of patch tests. *Br J Dermatol* 1999; 140: 102-5.
35. Ale SI, Maibach HI. Reproducibility of patch test results: a concurrent right-versus-left study using TRUE Test. *Contact Dermatitis* 2004; 50: 304-12.
36. Schiessl C, Wolber C, Strohal R. Reproducibility of patch tests: comparison of identical test allergens from different commercial sources. *Contact Dermatitis* 2004; 50: 27-30.
37. Ronchetti R, Jesenak M, Barberi S i wsp. Reproducibility of atopy patch tests with food and inhalant allergens. *J Biol Regul Homeost Agents* 2008; 22: 27-33.
38. Lazarov A, David M, Abraham D, Trattner A. Comparison of reactivity to allergens using the TRUE Test and IQ chamber system. *Contact Dermatitis* 2007; 56: 140-5.
39. Wang J, Godbold JH, Sampson HA. Correlation of serum allergy (IgE) tests performed by different assay systems. *J Allergy Clin Immunol* 2008; 121: 1219-24.
40. Greiner M, Pfeiffer D, Smith RD. Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests. *Prev Vet Med* 2000; 45: 23-41.
41. Akobeng AK. Understanding diagnostic tests 3: Receiver operating characteristic curves. *Acta Paediatrica* 2007; 96: 644-7.
42. Zweig MH, Campbell G. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clin Chem* 1993; 39: 561-77.
43. Pastorello EA, Incorvaia C, Ortolani C i wsp. Studies on the relationship between the level of specific IgE antibodies and the clinical expression of allergy: I. Definition of levels distinguishing patients with symptomatic from patients with asymptomatic allergy to common aeroallergens. *J Allergy Clin Immunol*. 1995; 96: 580-7.
44. Simoni M, Biavati P, Baldacci S i wsp. The Po River Delta epidemiological survey: reference values of total serum IgE levels in a normal population sample of North Italy (8-78 yrs). *Eur J Epidemiol* 2001; 17: 231-9.
45. Fernández C, Bevilacqua E, Fernández N i wsp. Asthma related to *Alternaria* sensitization: an analysis of skin-test and serum-specific IgE efficiency based on the bronchial provocation test. *Clin Exp Allergy* 2011; 41: 649-56.
46. Diaz-Vazquez C, Torregrosa-Bertet MJ, Carvajal-Urueña I i wsp. Accuracy of ImmunoCAP Rapid in the diagnosis of allergic sensitization in children between 1 and 14 years with recurrent wheezing: the IReNE study. *Pediatr Allergy Immunol* 2009; 20: 601-9.
47. Bernardini R, Pucci N, Azzari C i wsp. Sensitivity and specificity of different skin prick tests with latex extracts in pediatric patients with suspected natural rubber latex allergy--a cohort study. *Pediatr Allergy Immunol* 2008; 19: 315-8.
48. Dodig S, Richter D, Benko B i wsp. Cut-off values for total serum immunoglobulin E between non-atopic and atopic children in north-west Croatia. *Clin Chem Lab Med* 2006; 44: 639-47.
49. Carosso A, Bugiani M, Migliore E i wsp. Reference values of total serum IgE and their significance in the diagnosis of allergy in young European adults. *Int Arch Allergy Immunol* 2007; 142: 230-8.
50. Plebani M, Borghesan F, Basso D, Faggian D. Receiver-operating characteristic (ROC) curves: a fundamental tool for improving the clinical usefulness of in vitro IgE tests. *Allergy* 1996; 51: 407-11.
51. Rondón C, Canto G, Blanca M. Local allergic rhinitis: a new entity, characterization and further studies. *Curr Opin Allergy Clin Immunol* 2010; 10: 1-7.
52. Rondón C, Campo P, Herrera R i wsp. Nasal allergen provocation test with multiple aeroallergens detects polysensitization in local allergic rhinitis. *J Allergy Clin Immunol* 2011; 128: 1192-7.
53. Rondón C, Campo P, Togiás A i wsp. Local allergic rhinitis: concept, pathophysiology, and management. *J Allergy Clin Immunol* 2012; 129: 1460-7.

54. Poon AW, Goodman CS, Rubin RJ. In vitro and skin testing for allergy: comparable clinical utility and costs. *Am J Manag Care* 1998; 4: 969-85.
55. Hamilton RG, Biagini RE, Krieg EF. Diagnostic performance of Food and Drug Administration-cleared serologic assays for natural rubber latex-specific IgE antibody. The Multi-Center Latex Skin Testing Study Task Force. *J Allergy Clin Immunol* 1999; 103: 925-30.
56. Akkerdaas JH, Wensing M, Knulst AC i wsp. How accurate and safe is the diagnosis of hazelnut allergy by means of commercial skin prick test reagents? *Int Arch Allergy Immunol* 2003; 132: 132-40.
57. Niggemann B, Nilsson M, Freidrichs F. Pediatric allergy diagnosis in primary care is improved by in vitro allergen-specific IgE testing. *Pediatr Allergy Immunol* 2008; 19: 325-31.
58. Rajagopalan R, Anderson RT, Sarma S i wsp. The use of decision-analytical modelling in economic evaluation of patch testing in allergic contact dermatitis. *Pharmacoeconomics* 1998; 14: 79-95.
59. Rajagopalan R, Anderson RT, Sarma S i wsp. An economic evaluation of patch testing in the diagnosis and management of allergic contact dermatitis. *Am J Contact Dermat* 1998; 9: 149-54.
60. Śpiewak R. Patch Testing for Contact Allergy and Allergic Contact Dermatitis. *Open Allergy J* 2008; 1: 42-51.
61. Czarnobilska E, Obtułowicz K, Dyga W, Śpiewak R. The most important contact sensitizers in Polish children and adolescents with atopy and chronic recurrent eczema as detected with the expanded European Baseline Series. *Pediatr Allergy Immunol* 2011; 22: 252-6.
62. Gregorius A, Śpiewak R. Porównanie wyników testów płatkowych z zastosowaniem Europejskiej Serii Podstawowej, Polskiej Serii Podstawowej oraz rozszerzonej serii autorskiej w diagnostyce chorych z podejrzeniem alergii kontaktowej. *Alergoprofil* 2011; 7: 25-31.