

Ocena nowych metodologii badań przesiewowych



Royal College of
Obstetricians and
Gynaecologists

Ustalenie standardów postępowania
w celu poprawy zdrowia kobiet

Scientific Advisory Committee, Opinion Paper 20, kwiecień 2010

1. Wprowadzenie

Badania przesiewowe definiuje się jako badanie wykonywane u osób bez objawów w celu identyfikacji pacjentów z utajoną chorobą lub tych, u których istnieje duże prawdopodobieństwo przyszłego rozwoju choroby. Głównym celem badań przesiewowych jest umożliwienie interwencji poprawiających rokowanie dzięki leczeniu choroby we wcześniejszym stadium, bardziej poddającym się leczeniu lub zapobieganie jej rozwojowi. Badania przesiewowe mogą mieć różną formę, w tym: zbierania wywiadów, badania przedmiotowe, oznaczeń biochemicznych, obrazowania i badań cytologicznych. W ginekologii i położnictwie można znaleźć przykłady bardzo skutecznych i dobrze znanych badań przesiewowych, takich jak wykrywanie zespołu Downa oraz ocena cytologiczna wymazu z szyjki macicy. Skuteczne badania przesiewowe i interwencje mogłyby być pomocne w odniesieniu do wielu innych problemów w tej specjalności, takich jak urodzenie martwego płodu, przedwczesna niewydolność jajników i rak jajnika.

2. Ocena badań przesiewowych

Istnieje wiele parametrów stosowanych do ilościowej oceny różnych aspektów zdolności trafnej identyfikacji osób z grupy dużego ryzyka za pomocą testu przesiewowego. Poznanie znaczenia tych parametrów jest niezbędne dla zrozumienia możliwości wpływu danego testu na chorobę w populacjach lub u poszczególnych osób.

Czułość: Odsetek dodatnich wyników badania przesiewowego wśród wszystkich osób z daną chorobą (w populacji objętej oceną przesiewową)

Swoistość: Odsetek ujemnych wyników badania przesiewowego wśród wszystkich osób bez danej choroby (w populacji objętej oceną przesiewową)

Dodatnia wartość predykcyjna (PPV): Odsetek osób z daną chorobą wśród osób z dodatnim wynikiem badania przesiewowego

Ujemna wartość predykcyjna (NPV): Odsetek osób bez danej choroby wśród osób z ujemnym wynikiem badania przesiewowego

Wskaźnik wiarygodności dodatniego wyniku testu (positive likelihood ratio, LR+): Stosunek szansy choroby u osoby z dodatnim wynikiem badania przesiewowego do szansy choroby w populacji ogólnej

Wskaźnik wiarygodności ujemnego wyniku testu (negative likelihood ratio, LR-): Stosunek szansy choroby u osoby z ujemnym wynikiem badania przesiewowego do szansy choroby w populacji ogólnej

Być może najważniejszymi parametrami są: czułość, ponieważ wskazuje ona odsetek wszystkich przypadków chorobowych w populacji, którym moż-

na by zapobiec dzięki programowi badań przesiewowych (zakładając, że interwencja będzie skuteczna), oraz PPV, ponieważ parametr ten wskazuje, jak bardzo prawdopodobna jest choroba u osób z dodatnim wynikiem badania przesiewowego. Praktyczne znaczenie ma również odsetek dodatnich wyników badania przesiewowego, ponieważ określa on, jaka część populacji objętej badaniami przesiewowymi będzie wymagała dalszej diagnostyki.

Często stosowaną metodą charakteryzowania testów przesiewowych jest ocena krzywej operacyjno-charakterystycznej ROC. Jest to wykres, na którym na osi X zaznacza się 1 – swoistość (tj. odsetek wyników fałszywie dodatnich), a na osi Y – czułość. Każdy punkt na wykresie oznacza kombinację tych dwóch parametrów, przy założeniu pewnej wartości progowej definiującej dodatni wynik badania przesiewowego. Analiza pola pod krzywą ROC jest często wykorzystywanym sumarycznym parametrem statystycznym testów przesiewowych. Pole pod krzywą ROC oblicza się, porównując wszystkie możliwe kombinacje par osób z daną chorobą i bez niej. Wartość liczbowa pola pod krzywą ROC stanowi odsetek tych wszystkich porównań, w których u osoby z chorobą obliczona wartość ryzyka tej choroby jest większa niż u osoby bez choroby. W przypadku testu o idealnej wartości informacyjnej pole pod krzywą ROC wyniosłoby 1, a w przypadku testu całkowicie pozbawionego wartości informacyjnej – 0,5 (czyli wynik testu wypada nieprawidłowo równie często u osób zdrowych, jak u chorych). Ten parametr nie jest prosty

W imieniu Royal College of Obstetricians and Gynaecologists przygotował: Professor GSC Smith FRCOG, Cambridge i zrecenzowali:

Dr D Aitken, Glasgow; Professor DJ Murphy MRCOG; Dr RW Old, Warwick; Professor PW Soothill FRCOG. Głównym recenzentem z ramienia Scientific Advisory była Ms SM Quenby FRCOG.

W chwili publikacji przewodniczącym Scientific Advisory Committee był Professor S Thornton FRCOG, a wiceprzewodniczącym Professor R Anderson, FRCOG. Za ostateczną wersję odpowiada Scientific Advisory Committee RCOG.

RAMKA

Przypomnienie dotyczące pojęć ilorazu szans i wskaźnika wiarygodności

- Wskaźniki wiarygodności (likelihood ratio, LR) są sposobem modyfikacji szansy choroby na podstawie wyniku testu lub indywidualnej charakterystyki. Należy pamiętać o tym, że prawdopodobieństwo (probability) jest czymś innym niż szansa (odds).
- Szansa = $\frac{\text{prawdopodobieństwo}}{1 - \text{prawdopodobieństwo}}$. Odwrotnie, $\text{prawdopodobieństwo} = \frac{\text{szansa}}{1 + \text{szansa}}$.
- Kiedy dany stan występuje rzadko, szansa i prawdopodobieństwo są podobne. Jeżeli natomiast częstość występowania danego stanu jest duża, szansa i prawdopodobieństwo stają się różne.
- W przykładzie przytoczonym w tekście ryzyko choroby przed testem wynosi 20%, czyli $\text{prawdopodobieństwo} = 0,2$. Szansa choroby przed testem wynosi więc $0,2/0,8 = 0,25$.
- Szansa choroby po teście = szansa choroby przed testem \times wskaźnik wiarygodności (LR) = $0,25 \times 5,5 = 1,375$.
- Aby zamienić szansę choroby po teście z powrotem na prawdopodobieństwo, należy podzielić szansę przez (szansę + 1).
- $\text{Prawdopodobieństwo choroby po teście} = 1,375/2,375 = 0,58$. Przyjmując ryzyko przed testem wynoszące 20% oraz wskaźnik wiarygodności dodatniego wyniku testu wynoszący 5,5, ryzyko w przypadku dodatniego wyniku testu wynosi w przybliżeniu 60%.

konceptualnie (choć jest użyteczny w badaniach naukowych) i ma mniejsze bezpośrednie znaczenie niż pozostałe wyżej wymienione parametry. W wielu testach przesiewowych ocenia się zmienne ciągłe i posługiwanie się takim testem wymaga przyjęcia arbitralnej wartości progowej, powyżej której uważa się, że wynik testu przesiewowego jest dodatni. Ta przyjęta wartość progowa ryzyka określa również czułość oraz odsetek dodatnich wyników badania przesiewowego. Analiza krzywej ROC może być pomocna w wyznaczaniu tej wartości progowej. Wyznaczony próg zależy również od tego, jaki jest bilans konsekwencji niewykrycia istniejącej choroby oraz możliwych niepożądanych następstw (takich jak koszty, niepokój, ryzyko) dalszych badań u osób bez choroby. Zależnie od okoliczności, wyznaczając wartość progową, można więc kierować się minimalnym dopuszczalnym poziomem czułości lub maksymalną dopuszczalną częstością wyników fałszywie dodatnich.

3. Częstość występowania i zapadalność a badania przesiewowe

Interpretacja wszystkich powyższych danych zależy również od tego, jak czę-

sto dany stan występuje w populacji. Jeżeli występuje bardzo rzadko, nawet test o bardzo dużej wartości różnicującej może nie być przydatny klinicznie. Na przykład w badaniu przeprowadzonym w 2006 roku wykazano, że częstość występowania samoistnych urodzeń przedwczesnych przed ukończeniem 29 tygodni ciąży w populacji pierwsiastek, u których przeprowadzono przesiewowe badanie surowicy, wyniosła 0,24%.¹ W badaniu zastosowano test, dla którego wskaźnik wiarygodności dodatniego wyniku wyniósł 5,5. Ze względu na rzadkość tego zdarzenia prawdopodobieństwo jego wystąpienia u kobiet z dodatnim wynikiem badania przesiewowego wynosiło jednak tylko 1,3%. Gdyby natomiast częstość występowania jakiegoś stanu wynosiła 20%, to wówczas wskaźnik wiarygodności dodatniego wyniku testu wynoszący 5,5 identyfikowałby osoby, u których prawdopodobieństwo wystąpienia tego stanu w przybliżeniu wyniosłoby 60% (patrz ramka).

4. Ocena klinicznego zastosowania programu badań przesiewowych

Jedną z głównych cech charakteryzujących skuteczny program badań przesiewowych jest to, że składa się on

z dwóch odrębnych elementów. Pierwszym z nich jest zdolność identyfikacji osób z grupy dużego ryzyka, natomiast drugim możliwość zastosowania w tej grupie interwencji poprawiającej rokowanie. W idealnych warunkach te dwa elementy powinny być oceniane oddzielnie. Aby można było zaprojektować odpowiednią próbę kliniczną dotyczącą interwencji, muszą być znane właściwości testu jako metody badań przesiewowych. Wymaga to właściwie zaprojektowanych i przeanalizowanych badań dotyczących nowych metodologii diagnostycznych. Opracowano wytyczne Standards for the Reporting of Diagnostic Accuracy Studies (STARD), których celem jest wystandaryzowanie sposobu przedstawiania wyników oceny takich testów.²

W rzeczywistości w ginekologii i położnictwie istnieje wiele sytuacji, w których w badaniach naukowych nie uwzględniono istnienia tych dwóch odrębnych elementów badań przesiewowych. Ważnym przykładem jest rutynowe wykorzystywanie ultrasonografii w położnictwie. Przeprowadzono wiele prób klinicznych dotyczących tej kwestii. W metaanalizie nie wykazano wpływu rutynowego obrazowania, w tym dopplerowskich pomiarów prędkości przepływu krwi w krążeniu macicznie-łożyskowym, na umieralność w okresie okołoporodowym.^{3,4} W wielu badaniach nie stosowano jednak wystandaryzowanej interwencji u kobiet, u których uzyskano dodatni wynik oceny przesiewowej. Co więcej, te próby kliniczne zaprojektowano w sytuacji braku jednoznacznych informacji na temat właściwości testu jako metody przesiewowego wykrywania kobiet z grupy zwiększonego ryzyka zgonu w okresie okołoporodowym w populacji ogólnej. Nie jest więc jasne, czy w próbach klinicznych nie wykazano korzyści dlatego, że przesiewowa ocena nie pozwalała na identyfikację kobiet z grupy dużego ryzyka, z powodu niedostatecznej mocy statystycznej badań, czy też ze względu na niedostępność interwencji, która zmniejszyłaby ryzyko niepomysłnego rokowania u kobiet z dodatnim wynikiem oceny przesiewowej. Kwe-

stia ta została omówiona szczegółowo gdzie indziej.⁵

Ogólnie rzecz biorąc, randomizowane próby kliniczne powinny być przeprowadzane w grupach, w których uzyskano dodatni wynik badania przesiewowego, a nie analizowane na poziomie całej populacji objętej badaniami przesiewowymi. Nieprzestrzeganie tej zasady może spowodować, że skuteczne badanie przesiewowe i interwencja zostaną odrzucone jako nieskuteczne ze względu na niedostateczną moc statystyczną. Najlepiej ilustruje to hipotetyczny, ale prawdopodobny przykład. Rozważmy populację, w której zapadalność na cukrzycę ciążową wynosi 5%, możliwe jest leczenie tej cukrzyki, które zmniejsza chorobowość w okresie okołoporodowym z 4 do 1%, jak opisano w piśmiennictwie,⁶ chorobowość w okresie okołoporodowym wśród kobiet bez cukrzyki ciążowej wynosi 0,5% oraz, dla uproszczenia, dostępny jest przesiewowy test wykrywający cukrzycę ciążową ze swoistością i czułością równą 100%. Gdyby przeprowadzono randomizowaną kontrolowaną próbę kliniczną, przypisując kobiety do badań przesiewowych lub ich niewykonywania, chorobowość w okresie okołoporodowym wyniosłaby 0,525% w grupie interwencji oraz 0,675% w grupie kontrolnej. Aby uzyskać odpowiednią moc statystyczną umożliwiającą wykrycie tej różnicy, do badania trzeba by włączyć ponad 100 000 kobiet. Gdyby natomiast przeprowadzić randomizowaną kontrolowaną próbę kliniczną, porównując interwencję ze zwykłym leczeniem wśród kobiet, u których uzyskano dodatni wynik badania przesiewowego, to w celu wykrycia zmniejszenia ryzyka chorobowości w okresie okołoporodowym z 4% do 1% do badania trzeba by włączyć około 1250 kobiet z dodatnim wynikiem badania przesiewowego, a więc badaniami przesiewowymi należałoby objąć około 25 000 kobiet. Wy tłumaczeniem jest fakt, że moc statystyczna zwiększyła się przez wykluczenie większego odsetka niekorzystnych wyników leczenia, które były niezwiązane z cukrzycą ciążową, a więc program badań przesiewowych nie mógł mieć na nie wpływu.

Ważnym elementem oceny testów przesiewowych jest również ocena ich akceptacji, ponieważ nawet w przypadku doskonałej wartości predykcyjnej wpływ danego testu na rokowanie będzie niewielki, jeżeli większość kobiet nie będzie go akceptowała. Nastawienie konsumentów do nowych testów jest więc ważnym aspektem prób klinicznych dotyczących badań przesiewowych i interwencji. Niezbędne są wreszcie analizy ekonomiczne wszystkich nowych metod badań przesiewowych, ponieważ wprowadzenie nowych metod leczenia do praktyki klinicznej w ramach National Health Service (NHS) w Wielkiej Brytanii lub innych systemów opieki zdrowotnej gdzie indziej jest często uzależnione od kosztu przypadającego na jeden zyskany rok życia skorygowany o zmianę jego jakości.

5. Częste pułapki

W badaniach dotyczących testów przesiewowych często popełniane są pewne błędy. Po pierwsze, w wielu badaniach początkowo ocenia się testy przesiewowe w grupach pacjentów, u których ryzyko jest z natury duże (na przykład pacjenci trafiający do ośrodków trzeciego stopnia referencyjności). Jak omówiono wyżej, jedną z najważniejszych cech każdego testu jest jego PPV, która wykazuje silny związek z częstością występowania danego stanu. W początkowych opisach testu przesiewowego PPV jest więc często zawyżana. Po drugie, w wielu badaniach wykorzystuje się skomplikowane modele wykorzystujące wiele zmiennych w celu opracowania metody przewidywania ryzyka. Posługiwanie się modelami z dużą liczbą wskaźników predykcyjnych, zwłaszcza jeżeli wykorzystuje się je w odniesieniu do małej liczby przypadków, prowadzi jednak zwykle do nadmiernie optymistycznego szacowania wartości predykcyjnej. Modele powinny na ogół być weryfikowane w innej grupie uczestników niż ta, która posłużyła do opracowania modelu, zwłaszcza w badaniach obejmujących małą liczbę przypadków chorobowych i dużą liczbę wskaźników predykcyjnych. Po trzecie,

wczesne wykrycie choroby, w tym zwłaszcza nowotworu, prowadzi do wydłużenia czasu od rozpoznania do śmierci, nawet w przypadku braku skutecznej interwencji. Co więcej, rezultatem badań przesiewowych może być wykrycie choroby, która wiąże się z mniejszym prawdopodobieństwem zgonu lub wystąpienia objawów klinicznych, zanim nastąpi zgon z innych przyczyn. Badania przesiewowe mogą więc powodować pozorną poprawę rokowania mimo braku rzeczywistego rezultatu. Po czwarte, istnieje tendencja do rozważania badań przesiewowych wyłącznie na podstawie samego testu różnicującego. Badania przesiewowe powinny być rozważane tylko wtedy, gdy korzyści wynikające z wczesnego wykrycia choroby i interwencji jednoznacznie przeważają nad wadami programu badań przesiewowych (koszty, niepokój, chorobowość spowodowana procedurami diagnostycznymi wykonywanymi u osób z fałszywie dodatnim wynikiem badania przesiewowego i tak dalej). Problemem może być wreszcie niepełna identyfikacja przypadków chorobowych, która może prowadzić do mylących wniosków. Zławsza niewykrycie choroby u osób, u których uzyskano fałszywie ujemny wynik badania przesiewowego (tj. osób z daną chorobą, u których wynik badania przesiewowego wskazuje na jej brak), będzie prowadzić do nadmiernie optymistycznego szacowania efektywności testu przesiewowego.⁷

6. Opinia

Populacyjne badania przesiewowe są powszechnie stosowane w praktyce klinicznej oraz badaniach naukowych w ginekologii i położnictwie. Ocena wartości badań przesiewowych jest dość złożona metodologicznie i w tej specjalności często dokonuje się jej w suboptymalny sposób.

© 2010 Royal College of Obstetricians and Gynaecologists. Tłumaczenie i publikowanie artykułu Evaluating New Screening Methodologies, Opinion Paper 20 przez Medical Tribune Polska za zgodą RCOG. Jakielkolwiek kopiowanie w którymkolwiek języku w części lub w całości bez uprzedniego pisemnego zezwolenia wydawcy całkowicie zabronione.