

Tomasz Rzepiński

Making Diagnostic Strategies in Medical Practice with the Use of Bayes Theorem

Diametros nr 57, 39-60

2018

Artykuł został opracowany do udostępnienia w internecie przez Muzeum Historii Polski w ramach prac podejmowanych na rzecz zapewnienia otwartego, powszechnego i trwałego dostępu do polskiego dorobku naukowego i kulturalnego. Artykuł jest umieszczony w kolekcji cyfrowej bazhum.muzhp.pl, gromadzącej zawartość polskich czasopism humanistycznych i społecznych.

Tekst jest udostępniony do wykorzystania w ramach dozwolonego użytku.

Twierdzenie Bayesa w projektowaniu strategii diagnostycznych w medycynie

– Tomasz Rzepiński –

Abstrakt. W artykule porównane zostaną dwie metody wykorzystywane w projektowaniu strategii diagnostycznych. Pierwsza z nich to metoda określająca tzw. współczynnik predykcji. Druga opiera się na wykorzystaniu twierdzenia Bayesa. Zasadniczym celem artykułu jest zidentyfikowanie założeń epistemologicznych leżących u podstaw obu tych metod. Dla potrzeb realizacji tego celu rozważone zostaną przykładowe projekty jedno i wieloetapowej strategii diagnostycznej opracowanej z wykorzystaniem obu metod.

Słowa kluczowe: twierdzenie Bayesa, prawdopodobieństwo przedtestowe, strategie diagnostyczne, wartość predykcji, C. Howson, P. Urbach, G. Gigerenzer, U. Hoffrage, W. Jędrzychowski, T.J. Fagan.

Published online: 4 September 2018

1. Wstęp

Proces diagnozy lekarskiej stanowi przedmiot rozważań prowadzonych zarówno w płaszczyźnie społecznej, ekonomicznej jak i metodologicznej. Analizy metodologiczne mają na celu opracowanie strategii badawczych zwiększających efektywność procesu pozyskiwania informacji diagnostycznych. Diagnoza lekarska obejmuje trzy etapy: badanie podmiotowe (wywiad lekarski), badanie przedmiotowe (fizykalne) oraz badania dodatkowe, np. laboratoryjne i obrazowe¹. Dane pozyskiwane na dwóch pierwszych etapach diagnozy to zwykle dane o charakterze jakościowym². Pozwalają one na sformułowanie wstępnych hipotez stanowiących punkt wyjścia dla projektowania strategii diagnostycznych w etapie badań dodatkowych. Analiza metodologiczna ma umożliwić wybór takiej strategii diagnostycznej, która cechuje się największą efektywnością. Oznacza to, że wykonywane testy diagnostyczne powinny umożliwić pozyskiwanie informacji, które pozwolą dokonać trafnego rozpoznania choroby w jak najkrótszym czasie. Zadaniem analiz metodologicznych jest zatem opracowanie narzędzi formalnych umożliwiających ocenę efektywności testów diagnostycznych w taki sposób, aby można określić, które z nich dostarczą najbardziej istotnych dla procesu diagnozy informacji.

Tomasz Rzepiński
Instytut Filozofii UAM
ul. Szamarzewskiego 89c
60-568 Poznań
e-mail: rzepinskit@wp.pl

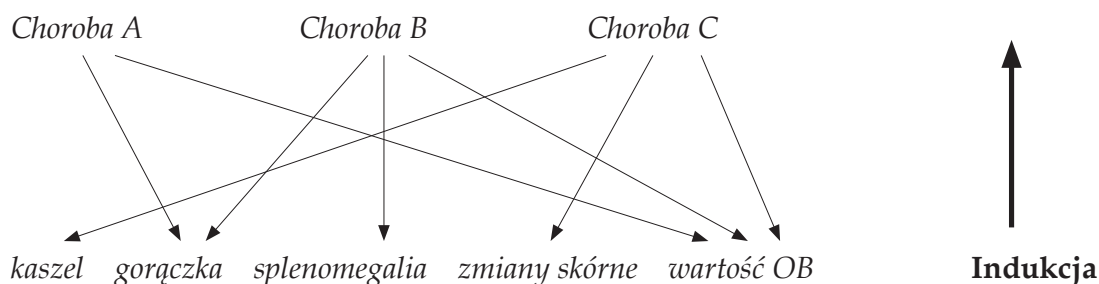
¹ Wulff (2005): 34–40.

² Jędrzychowski (1999): 179–182.

W artykule porównane zostaną dwie metody wykorzystywane w projektowaniu strategii diagnostycznych. Pierwsza z nich to metoda określająca tzw. współczynnik predykcji. Druga, opiera się na wykorzystaniu twierdzenia Bayesa. Zasadniczym celem artykułu jest zidentyfikowanie założeń epistemologicznych leżących u podstaw obu tych metod. Dla potrzeb realizacji tego celu rozważone zostaną przykładowe projekty jedno i wieloetapowej strategii diagnostycznej opracowanej z wykorzystaniem obu metod. Realizacja wskazanego celu wymaga przybliżenia podstawowych pojęć z zakresu metodologii medycyny, charakteryzujących testy diagnostyczne.

2. Wnioskowanie diagnostyczne

Testy diagnostyczne umożliwiają pozyskanie danych o cechach charakteryzujących pacjentów. W przypadku najbardziej informatywnych testów, zwykle biochemicznych, są to dane reprezentowane na skalach ilościowych³. Uzyskane w testach wyniki stanowią podstawę wnioskowań diagnostycznych, które mają na celu przyporządkowanie określonego pacjenta do określonej jednostki chorobowej. Wnioskowanie to jest wnioskowaniem indukcyjnym, którego przebieg można zilustrować następującym schematem.



Diagnosta, na podstawie pewnej skończonej ilości danych o cechach charakteryzujących pacjenta (ilościowych i jakościowych), wnioskuje o występowaniu choroby⁴.

Schemat 1

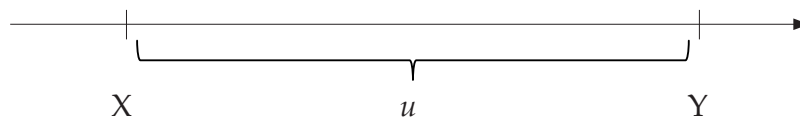
Problem, jak widać ze schematu 1, polega na tym, że te same wartości pewnych cech można przyporządkować do różnych chorób. Rozpoznanie choroby nie następuje zatem (zwykle) poprzez identyfikację wartości pojedynczej cechy lecz poprzez identyfikację określonego zestawu wartości cech, np. {podwyższona temperatura, ból, zmiany skórne} lub {obniżone stężenie pewnej substancji biochemicznej, podwyższona temperatura, tachykardia} itp.

Informacje o poszczególnych cechach diagnosta uzyskuje stopniowo. Niektóre z nich dostępne są już na etapie wywiadu, inne na etapie badania przedmiotowego, jeszcze inne dopiero w kolejnych fazach badań dodatkowych. Ze względu na stopniowe pozyskiwanie informacji, szczególnie istotny dla przebiegu diagnozy jest dobór kolejnych, odpowiednich testów diagnostycznych, takich, które w największym stopniu zwiększą lub zmniejszą

³ Rzepiński (2016).

⁴ Schemat za: Goodman (1999): 996.

prawdopodobieństwa wstępnych hipotez diagnostycznych, prowadząc ostatecznie do ich potwierdzenia lub wykluczenia. Jakie jednak wartości prawdopodobieństwa traktowane są jako wartości określające rozpoznanie lub wykluczenie choroby? Na to pytanie nie ma jednoznacznej odpowiedzi. Przeprowadzone badania pozwalają ustalić, że w procesie podejmowania decyzji lekarze przyjmują tzw. wartości progowe: próg wykluczenia i próg potwierdzenia, określane również odpowiednio jako próg testu i próg leczenia⁵.



X - próg wykluczenia, Y - próg leczenia, u - obszar niepewności.

Schemat 2

Wartości progów przyjmowane w różnych procesach diagnostycznych mogą się znacząco różnić w zależności od kosztów testów diagnostycznych i jatrogennego działania leków.

Projektowanie strategii diagnostycznych umożliwia ocenę przydatności informacji, które potencjalnie zostaną dopiero w danym teście pozyskane. Pozwala to oszacować, jaki zestaw dokonywanych po sobie testów pozwoliłby na przekroczenie progu potwierdzenia lub progu wykluczenia hipotez. Tworząc strategię diagnostyczną lekarz może posłużyć się jedną z dwóch metod: zastosować twierdzenie Bayesa lub obliczyć wartość predykcji.

3. Twierdzenie Bayesa

Twierdzenie Bayesa stanowiło przedmiot sporów prowadzonych zarówno w epistemologii jak i statystyce⁶. Pomimo formułowanych wobec niego zastrzeżeń, od lat 80-tych XX wieku jest ono coraz częściej wykorzystywane w diagnostyce medycznej. Przypomnijmy założenia leżące u jego podstaw.

Przyjmijmy, że dysponujemy zbiorem hipotez konkurencyjnych $\{H_1, \dots, H_n\}$, którym przypisujemy pewne wartości prawdopodobieństwa $P(H_1), \dots, P(H_n)$. Zakłada się, że hipotezy te wykluczają się, a suma ich prawdopodobieństw dopełnia się do jedności. Przyjęcie tych założeń pozwala zastosować twierdzenie Bayesa i określić z jego pomocą, jak zmieni się wartość prawdopodobieństwa poszczególnych hipotez po uzyskaniu nowych danych E , czyli obliczyć $P(H_i / E)$, dla $1 \leq i \leq n$. Ogólna postać tego twierdzenia jest następująca:

$$P(H_i / E) = \frac{P(H_i) \times P(E / H_i)}{P(H_1) \times P(E / H_1) + \dots + P(H_n) \times P(E / H_n)}$$

⁵ Jaeschke, Cook, Guyatt (1998): 158–159.

⁶ Obszerną charakterystykę stanowiska bayesianizmu czytelnik znajdzie w: Howson, Urbach (2006).

gdzie:

E – nowe dane

$P(H_i/E)$ – prawdopodobieństwo końcowe i -tej hipotezy po uzyskaniu danych E,

$P(H_i)$ – prawdopodobieństwo a priori hipotezy i -tej,

$P(E/H_i)$ – prawdopodobieństwo odwrotne, czyli prawdopodobieństwo uzyskania E na gruncie i -tej hipotezy.

Twierdzenie Bayesa może odgrywać istotną rolę w projektowaniu strategii diagnostycznych, pozwala bowiem ocenić wartość poznawczą danych, które diagnosta dopiero zamierza pozyskać w testach. W konsekwencji, zastosowanie tego twierdzenia w systemach wspomagania decyzji medycznych pozwoliłoby ograniczyć proces pozyskiwania informacji nieistotnych dla ostatecznego rozpoznania, zmniejszając jednocześnie koszty diagnostyki medycznej.

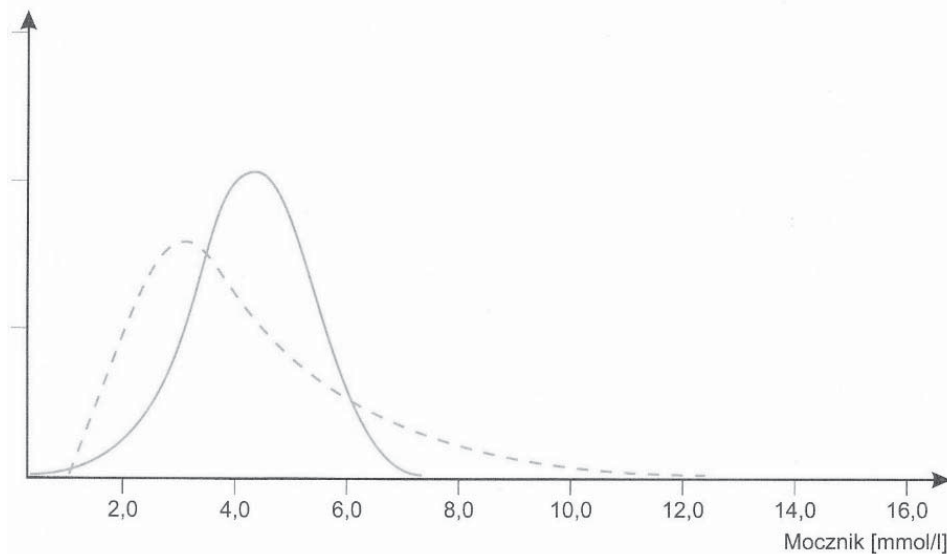
Możliwość wykorzystania twierdzenia Bayesa w projektowaniu strategii diagnostycznych budzi jednak wiele zastrzeżeń. W celu ich zilustrowania rozważmy prosty przykład wykorzystania tego twierdzenia dla oceny hipotez konkurencyjnych. Załóżmy, że dysponujemy urną, w której znajdują się kule czarne i białe w proporcji 1:3, jednak nie wiemy, których kul jest więcej. Na podstawie tego założenia można sformułować dwie hipotezy konkurencyjne: H_1 – kul białych jest więcej, H_2 – kul czarnych jest więcej. Hipotezom tym przypiszemy równe wartości prawdopodobieństwa, ponieważ przyjęte założenie nie preferuje żadnej z nich jako bardziej wiarygodnej, czyli $P(H_1) = 0,5$ oraz $P(H_2) = 0,5$. W dalszej kolejności zastanawiamy się, w jaki sposób zmieni się wartość prawdopodobieństwa obu tych hipotez po uzyskaniu informacji o wylosowaniu kuli czarnej. Aby zastosować twierdzenie Bayesa musimy znać wartości prawdopodobieństw odwrotnych. Pierwszym jest prawdopodobieństwo wylosowania kuli czarnej przy założeniu hipotezy H_1 , czyli: $P(E / H_1) = 0,25$. Drugim, prawdopodobieństwo wylosowania kuli czarnej przy założeniu hipotezy H_2 , czyli $P(E / H_2) = 0,75$. Dysponujemy zatem wszystkimi danymi, aby móc podstawić je do twierdzenia Bayesa. Dwukrotne zastosowanie tego twierdzenia wobec obu konkurencyjnych hipotez pozwoliłoby w dalszej kolejności w prosty sposób wykazać, że prawdopodobieństwa końcowe obu tych hipotez różnią się. Przykład ten nie został jednak przedstawiony po to, aby obliczać różnice końcowych wartości prawdopodobieństw konkurencyjnych hipotez, lecz po to aby ukazać jak bardzo silne założenia leżą u podstaw twierdzenia Bayesa.

Zauważmy, że w podanym przykładzie nie budzi wątpliwości posłużenie się statystyczną interpretacją prawdopodobieństwa w celu ustalenia prawdopodobieństwa a priori obu hipotez. Są to prawdopodobieństwa ustalane niezależnie od doświadczenia⁷. Po drugie, przyjęte założenia umożliwiają w sposób niebudzący wątpliwości obliczyć prawdopodobieństwa odwrotne obu hipotez. W diagnostyce medycznej spełnienie obu tych warunków, tj. określenie prawdopodobieństw a priori i określenie prawdopodobieństw odwrotnych związane jest z koniecznością dokonania wielu specyficznych rozstrzygnięć metodologicznych, które obniżają wiarygodność wniosków formułowanych z wykorzystaniem twierdzenia Bayesa.

⁷ Obszerne omówienie interpretacji prawdopodobieństwa czytelnik znajdzie m.in. w: Fine (1973): 85–117, Gillies (2000). Podstawowe różnice pomiędzy interpretacją statystyczną i częstościową w: Ajdukiewicz (1975).

4. Problemy epistemologiczne ustalenia zakresu wartości referencyjnych testu

Przyjmijmy, że diagnosta pragnie pozyskać informację o pewnej cesze charakteryzującej stan pacjenta. Załóżmy również, że wartości tej cechy mogą być reprezentowane na skali ilościowej. Jeżeli jej pomiar dokonany w teście diagnostycznym ma być wykorzystany w procesie rozpoznania choroby, to diagnosta musi dysponować wiedzą, jakie wartości tej cechy charakteryzują pacjentów zdrowych, a jakie pacjentów chorych. W zbiorze wartości, jakie przyjmuje dana cecha, muszą zatem być określone wartości graniczne. Ustalenie wartości granicznych nie jest jednak możliwe w zakresie nauk podstawowych. Badania prowadzone w biologii, biochemii, genetyce, immunologii i innych naukach podstawowych nie określają jakie wartości stężeń poszczególnych substancji biochemicznych są prawidłowe. Ustalenie takie może mieć wyłącznie charakter statystyczny⁸. Opiera się ono na określeniu rozkładu wartości danej cechy w populacji. Najczęściej rozkład wartości danej cechy w populacji jest rozkładem normalnym, reprezentowanym za pomocą krzywej Gaussa⁹. Idea określenia wartości granicznych jest zatem następująca. Przyjmuje się, że jeżeli dana cecha ma być użyteczna w procesie diagnozy, to wartości tej cechy w populacji osób zdrowych muszą różnić się od wartości w populacji osób chorych. Mamy zatem do czynienia z dwoma rozkładami: dla osób zdrowych i osób chorych. Gdyby test diagnostyczny miał charakter rozstrzygający jednoznacznie, to rozkłady te byłyby rozłączne. Tak jednak nie jest. Rozkłady te nakładają się na siebie w pewnym przedziale wartości, tak jak pokazano to w schemacie 3.



Częstość występowania stężeń mocznika u osób zdrowych (linia ciągła) i osób z chorobami nerek (linia przerywana). Schemat za: Naskalski (2005) s. 51.

Schemat 3

⁸ Dobrym przykładem jednego z pierwszych tego rodzaju ustaleń o charakterze statystycznym były badania prowadzone przez Biernackiego nad prędkością sedymentacji erytrocytów.

⁹ Oczywiście nie każda cecha jest reprezentowana za pomocą rozkładu normalnego. Patrz: Wulff, Göttsche (2005): 55.

Zauważmy, że w schemacie 3 większość osób, u których wartości stężeń mieszczą się w przedziale 4,0–6,0 to osoby zdrowe. Czy można zatem przyjąć właśnie te wartości stężeń jako wartości graniczne? Pytanie to podnosi ważny problem epistemologiczny dotyczący uzasadnienia przedziału wartości charakteryzujących osoby zdrowe, czyli tzw. przedziału wartości referencyjnych. Wyobraźmy sobie rozmowę prowadzoną pomiędzy statystykiem a dociekliwym studentem medycyny.

S. med.: Skąd wiemy, jaki jest przedział wartości referencyjnych?

Stat: Stąd, że przedział tych wartości charakteryzuje większość osób zdrowych.

S. med. Skąd wiemy, że większość wskazanych osób, to osoby zdrowe?

Stat: Stąd, że mieszczą się w przedziale wartości referencyjnych.

Uzasadnienie wartości granicznych podane przez statystyka ma zatem charakter błędnego koła. Na podstawie ustaleń o charakterze statystycznym określamy pojęcie zdrowia. Jednak samo uzasadnienie wartości granicznych zakłada już wiedzę o tym, które osoby są zdrowe¹⁰.

W praktyce medycznej wskazany problem rozwiązywany jest poprzez wykonanie testu referencyjnego, czyli testu traktowanego jako test ostatecznie rozstrzygający. To właśnie ten test ostatecznie określa, które osoby są, a które nie są zdrowe. Historycznie pierwszym opisem testu referencyjnego była tzw. triada Kocha, czyli postulaty, których spełnienie pozwalało zdiagnozować u danej osoby występowanie choroby zakaźnej. W zakresie współczesnej praktyki medycznej, przykładami bardziej znanych testów referencyjnych są: kolonoskopia w diagnostyce nowotworów jelita grubego lub koronarografia w diagnostyce niektórych chorób niedokrwiennych serca. Podkreślić należy, że przyjęcie testu referencyjnego nie rozwiązuje wskazanego problemu epistemologicznego. Uzasadnienie rozstrzygnięć testu referencyjnego jest bowiem nadal wystawione na wskazaną wcześniej krytykę. Przyjmując ostateczność tych rozstrzygnięć wystawiamy się na sceptyczny zarzut popełniania błędu *petitio principii*¹¹.

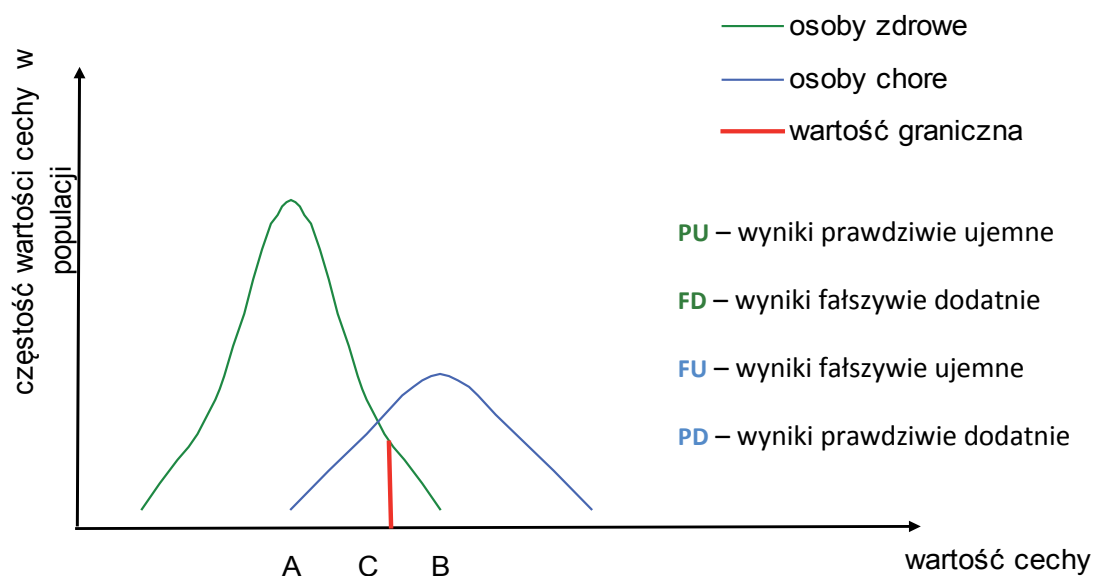
W świetle powyższych ustaleń zasadne jest pytanie o to, dlaczego w praktyce diagnostycznej nie ograniczyć się wyłącznie do stosowania testów referencyjnych. Odpowiedź wskazuje na problemy praktyczne wykorzystania tych testów. Kolonoskopia sprawia pacjentowi większy dyskomfort psychiczny niż badanie na krew utajoną. Koronarografia stwarza z kolei większe ryzyko uszczerbku na zdrowiu niż próba wysiłkowa itp. Testy referencyjne są ponadto zwykle bardziej kosztowne, czasochłonne, wymagają specjalistycznej aparatury.

¹⁰ Warto podkreślić, że jest to jeden z ważniejszych argumentów formułowanych przeciwko biostatystycznemu modelowi choroby. Porównaj: Sak (2013).

¹¹ Podkreślić należy, że w praktyce nie zawsze określony jest test referencyjny dla danego schorzenia. W takich sytuacjach dokonuje się wyłącznie przybliżonego oszacowania parametrów testu. Omówienie tych zagadnień wykracza poza ramy artykułu. Czytelników zainteresowanych odsyłam do: Joseph, Gyorkos, Coupal (1995); Enøe, Georgiadis, Johnson (2000); Hui, Zhou (1998).

5. Parametry oceny testu diagnostycznego

Wyniki testów potwierdzających chorobę zwyczajowo określa się mianem wyników dodatnich lub pozytywnych. Z kolei wyniki wykluczające chorobę - ujemnych lub negatywnych. Ocena wartości poznawczej testów diagnostycznych dokonywana jest z wykorzystaniem kilku parametrów. Wskażę na razie dwa podstawowe. Pierwszym jest czułość diagnostyczna, drugim swoistość diagnostyczna. Czułość diagnostyczna (Cz.d) to zdolność testu do potwierdzenia choroby. Z kolei swoistość diagnostyczna (Sw.d) to zdolność testu do wykluczenia choroby¹². Oba te parametry określane są w sposób ilościowy. W celu ich sprecyzowania rozważmy następujący schemat ilustrujący wyniki testu diagnostycznego T w pewnej populacji osób.



Osoby zdrowe to osoby rozpoznane jako zdrowe na podstawie testu referencyjnego, analogicznie chore. Linia czerwona w punkcie C określa wartość graniczną rozróżniającą w teście T osoby zdrowe i chore.

Schemat 4

Zauważmy, że w przedziale wartości od A do B rozważany test diagnostyczny nie ma charakteru rozstrzygającego. Pacjenci zdrowi i chorzy są nieodróżnialni na podstawie wartości z tego przedziału. Występowanie obszaru niejednoznacznych rozstrzygnięć zmniejsza wartość poznawczą testu. Wśród wyników dodatnich należy bowiem wyróżnić te, które są prawdziwie dodatnie (PD) i te, które są fałszywie dodatnie (FD). Analogicznie wśród ujemnych: prawdziwie ujemne (PU) i fałszywie ujemne (FU). Pamiętać należy, że fałszywość oraz prawdziwość wyników jest zrelatywizowana do arbitralnych ustaleń testu referencyjnego. Wskazana klasyfikacja wyników wykorzystywana jest w charakterystyce czułości i swoistości diagnostycznej testu.

¹² Jędrzychowski (2002): 117-120.

Przyjmijmy, że n oznacza licznosc zbioru. Wówczas:

$$\text{Cz.d} = \frac{n(PD)}{n(PD + FU)}$$

$$\text{Sw.d} = \frac{n(PU)}{n(PU + FD)}$$

Rozważmy prosty przykład ilustrujący te ustalenia. Załóżmy, że chcemy ocenić czułość i swoistość diagnostyczną testu T_α na podstawie rozstrzygnięć z testu referencyjnego zilustrowanych za pomocą następującej tabeli.

Tabela 1

	Choroba obecna	Choroba nieobecna
T_α^+	PD = 42	FD = 26
T_α^-	FU = 10	PU = 55

T_α^+ - wynik dodatni, T_α^- - wynik ujemny. Rozróżnienie na chorobę obecną i nieobecną na podstawie testu referencyjnego.

$$\text{Cz.d} \text{ wynosi: } \frac{42}{42 + 10} = 80,77\%$$

$$\text{Odpowiednio Sw.d} \text{ wynosi: } \frac{55}{55 + 26} = 67,9\%$$

Na podstawie tych danych lekarz może zatem stwierdzić, że wśród 100 pacjentów chorych, pozytywny wynik testu (T_α^+) pozwoli przypuszczalnie wykryć schorzenie u około 81 z nich, a u 19 choroba nie zostanie rozpoznana. Z kolei na 100 pacjentów zdrowych negatywny wynik testu (T_α^-) pozwoli wykluczyć chorobę u około 68 z nich, a 32 zdrowych będzie zbędnie podlegało dalszej diagnostyce.

Projektując strategię diagnostyczną lekarz sięga do publikacji opisujących podstawowe parametry testów diagnostycznych. Informacje o wartościach czułości i swoistości diagnostycznej testu stanowią ważną część wiedzy wpływającą na proces wyboru testu w określonej sytuacji diagnostycznej. Wartości te mogą być jednak ustalane w badaniach populacji, w których częstość występowania danej choroby różni się od częstości występowania choroby w populacji, z którą w swej praktyce ma do czynienia lekarz. Przykładowo wiadomo, że testy biochemiczne stosowane współcześnie w diagnostyce zawału serca posiadają bardzo wysokie wartości czułości i swoistości diagnostycznej. Jednak choroba niedokrwienna serca jest chorobą cywilizacyjną, ze zwiększoną częstością występowania w społeczeństwach konsumpcyjnych. W tej sytuacji, dla projektowania strategii diagnostycznych i podejmowania decyzji ważne jest ustalenie, czy oba parametry testu zależą od częstości występowania choroby w populacji. Odnosząc się do założeń z tabeli 1 łatwo jest się przekonać, że parametry te nie ulegną zmianie, przy zmianie częstości występowania choroby. Przykładowo, dla tabeli 1 przyjęcie założenia o dwukrotnym zwiększeniu częstości występowania choroby spowoduje dwukrotne zwiększenie zarówno ilości wyników PD (2×42) jak i wyników FU (2×10). Proporcja wyników charakteryzujących czułość i swoistość diagnostyczną pozostanie jednak stała. Niezależność czułości i swoistości diagnostycznej od częstości występowania choroby sprawia, że oba te parametry

uzyskują uniwersalny charakter. Dostarczają takich informacji o teście, które mogą być wykorzystane wobec różnych populacji pacjentów¹³.

Scharakteryzowanie czułości i swoistości diagnostycznej pozwala zastanowić się, jaką funkcję pełnią oba te parametry w projektowaniu strategii diagnostycznych wykorzystujących twierdzenie Bayesa.

6. Twierdzenie Bayesa a parametry oceny testu diagnostycznego

Przyjmijmy, że diagnosta w pewnej konkretnej sytuacji diagnostycznej formułuje dwie hipotezy: H^+ - choroba D występuje u pacjenta P, H^- - choroba D nie występuje u pacjenta P. Diagnosta projektując strategię diagnostyczną rozważa przeprowadzenie testu T_α . Załóżmy, że przyjmuje próg potwierdzenia hipotezy H^+ na poziomie 90% a próg jej wykluczenia na poziomie 10%¹⁴. Następnie zastanawia się w jakim stopniu pozytywny wynik testu (T_α^+) zmieni wstępne wartości prawdopodobieństwa obu hipotez oraz jak prawdopodobieństwa tych hipotez ulegną zmianie w efekcie uzyskania wyniku negatywnego (T_α^-). Pragnie zatem określić wartości: $P(H^+ / T_\alpha^+)$, $P(H^- / T_\alpha^-)$, $P(H^+ / T_\alpha^-)$, $P(H^- / T_\alpha^+)$, tak aby oszacować, czy warto jest test T_α przeprowadzać. Stara się zatem w istocie ustalić, w jakim stopniu dany test zbliży wartość prawdopodobieństwa hipotezy do jednego z założonych progów. Dla uproszczenia rozważmy wyłącznie wartość $P(H^+ / T_\alpha^+)$. Określając tę wartość diagnosta może posłużyć się twierdzeniem Bayesa odpowiednio je modyfikując:

$$P(H^+ / T_\alpha^+) = \frac{P(H^+) \times P(T_\alpha^+ / H^+)}{P(H^+) \times P(T_\alpha^+ / H^+) + P(H^-) \times P(T_\alpha^+ / H^-)}$$

Zgodnie z dokonanymi wcześniej uwagami wykorzystanie twierdzenia Bayesa wiąże się z dwoma problemami. Pierwszym jest obliczenie prawdopodobieństw odwrotnych, czyli $P(T_\alpha^+ / H^+)$ oraz $P(T_\alpha^+ / H^-)$. Dla potrzeb rozstrzygnięcia tej kwestii można odwołać się do statystycznej interpretacji pojęcia prawdopodobieństwa¹⁵. Zgodnie z nią:

$$P(A / B) = \frac{n(A \cap B)}{n(B)}$$

Zatem:

$$P(T_\alpha^+ / H^+) = \frac{n(T_\alpha^+ \cap H^+)}{n(H^+)}$$

Łatwo stwierdzić na podstawie schematu 4, że $n(T_\alpha^+ \cap H^+) = PD$, z kolei $n(H^+) = PD + FU$. Zatem:

$$P(T_\alpha^+ / H^+) = \frac{n(PD)}{n(PD + FU)}, \text{ czyli jest to czułość diagnostyczna.}$$

¹³ Odrębnym problemem jest natomiast zmiana wartości obu tych parametrów w zależności od stopnia rozwoju choroby. Porównaj: Joseph, Gyorkos, Coupal (1995).

¹⁴ Ponieważ prawdopodobieństwa hipotez dopełniają się do jedności, zatem potwierdzenie H^+ jest równoznaczne z wykluczeniem H^- i na odwrót.

¹⁵ Ajdukiewicz (1975).

Prowadząc podobne rozumowanie można wykazać z kolei, że:

$$P(T_{\alpha}^{+}/H) = 1 - \text{swoistość diagnostyczna.}$$

Zatem po uwzględnieniu charakterystyki testów diagnostycznych struktura twierdzenia Bayesa dla rozważanego przykładu ma postać:

$$P(H^{+} / T_{\alpha}^{+}) = \frac{P(H^{+}) \times Cz.d}{P(H^{+}) \times Cz.d + P(H^{-}) \times (1 - Sw.d)}$$

Dokonane przekształcenie wydawać się może wyłącznie ciekawostką ilustrującą dostosowanie formy twierdzenia Bayesa do specyfiki procedur diagnostycznych. W kontekście uzyskanego przekształcenia warto jednak zwrócić uwagę na zagadnienie, które posiada istotne konsekwencje epistemologiczne. W celu jego omówienia powróćmy do problemu ustalania wartości referencyjnych.

W schemacie 4 wartość graniczna została ustalona w punkcie C. Zauważmy jednak, że ustalenie takie ma całkowicie arbitralny charakter. Fakt, że dwa rozkłady przecinają się w punkcie reprezentującym wartość C nie uzasadnia przecież ustalenia wartości granicznej właśnie w tym miejscu. Zastanówmy się zatem, jak zmienią się parametry testu diagnostycznego w sytuacji, gdy wartość graniczną przesuniemy do punktu A.

Przesunięcie wartości granicznej w schemacie 4 do punktu A wyeliminuje wyniki fałszywie ujemne. Wówczas czułość diagnostyczna wzrośnie do 100%, ponieważ: $\frac{n(PD)}{n(PD + 0)}$.

Wzrost czułości diagnostycznej spowodowany przesunięciem wartości granicznej spowoduje jednak zmniejszenie swoistości diagnostycznej, ponieważ zwiększy się ilość wyników FD. Analogicznie możemy zwiększyć swoistość diagnostyczną do 100% poprzez przesunięcie wartości granicznej do punktu B, kosztem zmniejszenia czułości diagnostycznej¹⁶. Widać zatem wyraźnie, że ustalenie wartości granicznej ma charakter kompromisu, motywowanego z jednej strony dążeniem do zminimalizowania strat społecznych spowodowanych nierozpoznanie choroby, z drugiej zaś chęcią zmniejszenia kosztów zbędnej diagnostyki medycznej i angażowania personelu w diagnostykę pacjentów zdrowych.

Jeżeli jednak ustalenie wartości granicznych ma charakter arbitralny, to w znaczący sposób zmniejsza to obiektywność wniosków formułowanych na podstawie twierdzenia Bayesa. Wartości czułości i swoistości diagnostycznej zależą bowiem od tego, w którym miejscu określona zostanie wartość graniczna. Twierdzenie Bayesa pozwala zatem określić jak zmieni się wartość prawdopodobieństwa hipotez diagnostycznych, o ile dokonamy już wcześniej pragmatycznych rozstrzygnięć dotyczących zakresu przedziałów referencyjnych.

7. Problem wyznaczania prawdopodobieństw a priori hipotez diagnostycznych

Krytyka twierdzenia Bayesa podnosi na pierwszym miejscu sposób wyznaczania wartości prawdopodobieństw a priori. Uznaje się mianowicie, że określenie wartości prawdopodobieństwa a priori hipotez formułowanych wobec nietrywialnych sytuacji

¹⁶ Jędrychowski (2002): 114.

problemowych wymaga przyjęcia interpretacji personalistycznej. W przykładzie dotyczącym losowania kul, prawdopodobieństwo a priori mogło być określone na gruncie interpretacji statystycznej, ponieważ opisana tam sytuacja problemowa miała trywialny charakter. Dotyczyła skończonego zbioru zdarzeń. W jaki jednak sposób określić wartość prawdopodobieństwa a priori hipotez w nietrywialnych sytuacjach, jakimi są sytuacje praktyki diagnostycznej?

Przede wszystkim należy zaznaczyć, że wartość prawdopodobieństwa a priori nie jest rozumiana w literaturze medycznej jako wartość ustalana niezależnie od doświadczenia, lecz jako początkowa wartość prawdopodobieństwa [pretest probability] przypisywana wstępnie hipotezom diagnostycznym¹⁷. Termin „początkowa” nie oznacza niezależnej od doświadczenia, niezależnej od względnej częstości wcześniejszych zdarzeń.

Formułując hipotezy diagnostyczne na podstawie badania podmiotowego i przedmiotowego, lekarz odwołuje się do swoich dotychczasowych doświadczeń obejmujących zbliżone przypadki występowania symptomów u pacjentów badanych przez niego wcześniej. Kieruje się zatem regułą określającą prawdopodobieństwo w sposób częstościowy¹⁸. Problemem pojawiającym się w przypadku wyznaczania wartości prawdopodobieństwa w oparciu o indywidualne doświadczenia lekarzy jest zróżnicowanie tych doświadczeń. Posiada to istotne konsekwencje dla przebiegu praktyki. Lekarze praktykują w obszarach o różnych częstościach występowania choroby. Praktykujący w populacjach, w których choroba występuje często są bardziej skłonni przypisywać wysokie prawdopodobieństwo hipotezom identyfikującym tę chorobę, aniżeli lekarze praktykujący w populacjach, w których to schorzenie jest rzadsze.

W celu zmniejszenia różnic w wartościach prawdopodobieństwa wstępnego przypisywanych przez różnych diagnostów publikowane są dane dotyczące częstości występowania choroby w określonych subpopulacjach. Dane takie traktowane są jako zobiektywizowane wartości prawdopodobieństwa częstościowego¹⁹.

Wartość prawdopodobieństwa wstępnego interpretowanego jako prawdopodobieństwo częstościowe można również określić w badaniach epidemiologicznych obliczając tzw. współczynnik chorobowości, na podstawie danych z tabeli 1. Wyraża on stosunek ilości wszystkich pacjentów chorych do ilości wszystkich pacjentów poddanych testowi, czyli:

$$Wsp.ch = \frac{n(PD + FU)}{n(PD + PU + FD + FU)}$$

Współczynnik chorobowości jest bardzo precyzyjnym parametrem o wysokiej wartości poznawczej dla procesu projektowania strategii diagnostycznej. Zauważmy, że nie wyraża on częstości występowania choroby w całej populacji. Nie określa on np. częstości występowania nowotworu płuc w całej populacji Dolnego Śląska. Jest on zawężony wyłącznie do tej populacji, która została poddana w przeszłości testom, czyli tym osobom, u których wykonanie testu wydawało się diagnostom uzasadnione na podstawie

¹⁷ Paulo, Mendes, Vizinho et al. (2004); Steurer (2009).

¹⁸ „Już tak prosta informacja, jak wiek pacjenta, pozwala precyzyjniej określić prawdopodobieństwo [...] występowania astmy i niewydolności serca” – Jaeschke, Cook, Guyatt (1998): 158.

¹⁹ Paulo, Mendes, Vizinho et al. (2004).

wyników wywiadu i badania przedmiotowego. Gdybyśmy rozpatrywali całą populację Dolnego Śląska, to ujęlibyśmy również dzieci i ludzi bardzo młodych, u których nowotwór płuc występuje zapewne rzadko i które nigdy do testu nie były kwalifikowane. Jeżeli zatem, podejmując diagnozę wobec konkretnego pacjenta, lekarz na podstawie przeprowadzonego u niego badania podmiotowego i przedmiotowego uznaje, że jest uzasadnione podejrzenie, że pacjent ten cierpi na nowotwór płuc, to oznacza to, że pacjent ten mieści się we wzorcu epidemiologicznym pacjentów uwzględnionych za pomocą współczynnika chorobowości. Stąd właśnie precyzja tego współczynnika jako parametru reprezentującego w twierdzeniu Bayesa wstępną wartość prawdopodobieństwa.

8. Projektowanie strategii diagnostycznych

Ustalenie wartości prawdopodobieństw odwrotnych oraz rozstrzygnięcie sposobu ustalania wartości prawdopodobieństw wstępnych pozwala zastosować twierdzenie Bayesa w ocenie użyteczności informacji, które diagnosta dopiero zamierza uzyskać w teście. Rozważmy przykład oparty na danych z tabeli 1. Przyjmijmy, że diagnosta pragnie ustalić w jakim stopniu pozytywny wynik testu (T_{α}^{+}) otrzymany u konkretnego pacjenta zmieni wartość prawdopodobieństwa hipotezy (H^{+}) o występowaniu u niego choroby D, przyjmując wskazane wcześniej progi potwierdzenia i wykluczenia hipotezy. Stosuje zatem twierdzenie Bayesa o podanej wcześniej postaci:

$$P(H^{+} / T_{\alpha}^{+}) = \frac{P(H^{+}) \times Cz.d}{P(H^{+}) \times Cz.d + P(H^{-}) \times (1 - Sw.d)}$$

Zgodnie z ustaleniami z paragrafu 4 dla danych z tabeli 1 czułość diagnostyczna testu T_{α} wynosiła 81%, natomiast swoistość diagnostyczna 63%. Prawdopodobieństwo wstępne $P(H^{+})$ obliczone jako współczynnik chorobowości wynosi:

$$P(H^{+}) = Wsp.ch = \frac{n(PD + FU)}{n(PD + PU + FD + FU)} = \frac{42 + 10}{42 + 51 + 30 + 10} = 39\%$$

Ponieważ konkurencyjne hipotezy diagnostyczne dopełniają się do jedności, zatem:

$$P(H^{-}) = 1 - P(H^{+}) = 1 - 0,39 = 61\%$$

W efekcie podstawienia danych do twierdzenia Bayesa uzyskujemy:

$$P(H^{+} / T_{\alpha}^{+}) = \frac{0,39 \times 0,81}{0,39 \times 0,81 + 0,61 \times (1 - 0,68)} = 62\%$$

Zastosowanie twierdzenia Bayesa pozwala zatem diagnoście stwierdzić, że uzyskanie pozytywnego wyniku testu u pewnego pacjenta zwiększy prawdopodobieństwo hipotezy o występowaniu danej choroby u tego pacjenta z początkowej wartości 39% do wartości 62%. Nie zostaje jednak przekroczony żaden z przyjętych progów decyzyjnych. Jeżeli zatem taka zmiana jest dla diagnosty niesatysfakcjonująca, zbyt mała, to może on poszukiwać testu diagnostycznego o innych parametrach, który, na kolejnym etapie diagnozy, w większym stopniu wpłynie na zmianę wstępnej wartości prawdo-

podobieństwa hipotezy²⁰. Jakimi jednak wartościami parametrów powinien cechować się test o wysokiej mocy potwierdzającej hipotezę H^+ ?

Przypomnijmy, że zdolność testu do potwierdzenia hipotezy określana jest jako czułość diagnostyczna. Można by zatem sądzić, że test o wysokiej czułości diagnostycznej posiada wysoką moc potwierdzania hipotezy H^+ . Spójrzmy zatem na schemat 4 i zastanówmy się, jak zmieni się stan epistemiczny diagnosty wobec hipotezy H^+ w przypadku testu o 100% czułości diagnostycznej. Jeżeli czułość diagnostyczna testu wynosi 100%, to wartość graniczna w schemacie 4 jest określona w punkcie A. Widać jednak wyraźnie, że w takiej sytuacji uzyskanie pozytywnego wyniku testu (czyli wynik na prawo od A) nie dostarcza lekarzowi pewności rozpoznania choroby u danego pacjenta. Wśród wyników pozytywnych występują bowiem zarówno wyniki prawdziwe, jak i fałszywe. Test o wysokiej czułości diagnostycznej nie jest zatem testem, którego pozytywny wynik dostarcza pewności, że u konkretnego pacjenta występuje choroba. Jest to parametr, który gwarantuje, że pozytywne wyniki uzyskane z takiego testu wychwycą wszystkich pacjentów chorych. Innymi słowy diagnosta zyskuje gwarancję, że każdy chory poddany temu testowi będzie miał pozytywny wynik, ale nie jest tak, że każdy pozytywny wynik identyfikuje chorego. Jeżeli raz jeszcze spojrzymy na schemat 4, to zauważymy, że test o wysokiej czułości diagnostycznej dostarcza diagnoście pewności wykluczenia hipotezy H^+ . Negatywne wyniki testu o 100% czułości diagnostycznej obejmują bowiem wyłącznie wyniki prawdziwe w odniesieniu do testu referencyjnego. Analogiczne ustalenia można dokonać w odniesieniu do zmiany stanu epistemicznego diagnosty na podstawie testu o 100% swoistości diagnostycznej, w przypadku przesunięcia w schemacie 4 wartości granicznej do punktu B. Test taki dostarcza pewności potwierdzenia hipotezy H^+ wobec testu referencyjnego, nie nadaje się natomiast dla wykluczenia tej hipotezy.

Dokonane ustalenia są niezwykle istotne z perspektywy procesu prawidłowego projektowania strategii diagnostycznych. Odmienne wartości parametrów muszą charakteryzować testy wykorzystywane w celu potwierdzania hipotez, inne natomiast muszą posiadać testy używane dla potrzeb ich wykluczenia. Spostrzeżenia te są szczególnie istotne w przypadku projektowania wieloetapowej strategii diagnostycznej. Należy bowiem wówczas wyraźnie określić, jakie cele poznawcze realizowane będą na kolejnych etapach procesu diagnozy. Przykładowo, w pierwszym etapie programu diagnostyki przesiewowej musi być zastosowany test o wysokiej czułości, bowiem celem na tym etapie programu jest wychwycenie wszystkich chorych we wczesnej fazie choroby, nawet kosztem uzyskania wielu wyników fałszywie dodatnich. W kolejnym etapie programu

²⁰ W tym miejscu konieczne jest istotne dopowiedzenie. Otóż poza wartością wstępnego prawdopodobieństwa mającą wpływ na wartość predykcijną testu, kolejnymi parametrami, które wpływają na tę wartość, są: czułość i swoistość diagnostyczna. Należy pamiętać, że podawane w literaturze przedmiotu wartości czułości i swoistości diagnostycznej testu są określone dla „typowych” przypadków przebiegu choroby, na uśrednionym poziomie jej zaawansowania. Dalszy rozwój choroby może wpłynąć na to, że wartości czułości i swoistości diagnostycznej ulegną zmianie. W efekcie możemy mieć do czynienia z sytuacją, w której powtórzenie tego samego testu po jakimś czasie dostarczyć może nie tylko bardziej lub mniej wiarygodnych informacji, ale przede wszystkim informacji odmiennych od tych, które uzyskane zostały przy pierwszym wykonaniu testu. Na problem niejednoznaczności w ustalaniu wartości czułości i swoistości diagnostycznej zwracają uwagę m.in. Moons, van Es, Deckers et al. (1997) oraz Winkler, Smith (2004).

stosuje się natomiast test o wysokiej swoistości, dla ustalenia, którzy pacjenci wskazani w pierwszym teście jako potencjalnie chorzy są rzeczywiście chorzy.

Opis zastosowania twierdzenia Bayesa w projektowaniu strategii diagnostycznych pozostawałby niepełny bez omówienia jeszcze jednej kwestii. Otóż poza wymienionymi parametrami, tj. czułością i swoistością diagnostyczną oraz współczynnikiem chorobowości, można wskazać na jeszcze jeden ważny parametr charakteryzujący test diagnostyczny. Jest nim współczynnik predykcji. Parametr ten określa prawdopodobieństwo występowania choroby u osoby z danym wynikiem testu. Wyróżniamy predykcję dodatnią (Pred.⁺) i predykcję ujemną (Pred.⁻)²¹. Dla potrzeb prowadzonych rozważań wystarczy scharakteryzować pierwszą z nich. Predykcja dodatnia określa stosunek ilości wyników PD do ilości wszystkich wyników dodatnich (schemat 4). Innymi słowy jest to aposteriorycznie rozumiane prawdopodobieństwo tego, że wśród wyników dodatnich uzyskamy wynik prawdziwie dodatni, czyli:

$$\text{Pr } ed.^+ = \frac{n(PD)}{n(PD + FD)}$$

Zauważmy, że w istocie parametr ten realizuje dokładnie ten sam cel poznawczy, jaki realizuje twierdzenie Bayesa²². Spostrzeżenie to łatwo potwierdzić podstawiając odpowiednie dane z tabeli 1:

$$\text{Pr } ed.(T_{\alpha}^+) = \frac{42}{42 + 26} = 62\%$$

Jaki jest zatem sens posługiwania się twierdzeniem Bayesa, jeżeli ten sam wynik można uzyskać na podstawie współczynnika wartości predykcji, posługując się interpretacją częstościową i nie narażając się na zarzut personalizmu? Odpowiedź jest prosta. Otóż różnica pomiędzy wynikami uzyskanymi za pomocą współczynnika predykcji i tymi uzyskanymi za pomocą twierdzenia Bayesa ujawnia się dopiero w przypadku projektowania wieloetapowych strategii diagnostycznych²³.

9. Twierdzenie Bayesa i współczynnik predykcji w projektowaniu wieloetapowej strategii diagnostycznej

Załóżmy, że diagnosta zastosował w projekcie strategii diagnostycznej twierdzenie Bayesa lub współczynnik wartości predykcji, uzyskując opisany powyżej wynik. Stwierdził zatem,

²¹ Jędrychowski (1999): 129.

²² Wulff, Gøtzsche (2005): 101; Jędrychowski (1999): 131.

²³ Równoważność obliczeń dokonywanych za pomocą twierdzenia Bayesa i współczynnika predykcji stanowiła podstawę dla analiz mających na celu określenie, w jaki sposób prowadzone są wnioski diagnostyczne – Gigerenzer, Hoffrage (1995): 687. Autorzy pokazali, że diagności znacznie więcej problemów mieli z poprawnym zastosowaniem twierdzenia Bayesa niż z ustaleniem wartości współczynnika predykcji. Zastosowanie twierdzenia Bayesa wymaga znajomości pojęć z zakresu biostatystyki, podczas gdy obliczenie współczynnika predykcji opiera się na bardziej naturalnym, z perspektywy ewolucyjnej, częstościowym rozumieniu prawdopodobieństwa. Wnioski autorów wskazywały na konieczność podniesienia poziomu edukacji w zakresie biostatystyki – ibidem: 701. Podkreślić należy, że praca Gigerenzera i Hoffrage'a dotyczyła wyłącznie jednoetapowego procesu diagnozy medycznej. Nie wykazywano zatem różnicowania wyników uzyskiwanych za pomocą obu metod w procesie projektowania strategii wieloetapowych.

że po otrzymaniu pozytywnego wyniku testu w pierwszym etapie diagnozy, wstępna wartość prawdopodobieństwa hipotezy H^+ ulegnie zwiększeniu z 39% do 62%. Pragnąc w większym stopniu potwierdzić tę hipotezę diagnosta musi w projekcie kolejnego etapu swojej strategii uwzględnić test o wyższej swoistości diagnostycznej (paragraf 7). Przyjmijmy, że test ten (T_β) scharakteryzowany jest, wobec testu referencyjnego, w następujący sposób.

Tabela 2

	Choroba obecna	Choroba nieobecna
T_β^+	PD = 98	FD = 18
T_β^-	FU = 23	PU = 171

Na podstawie danych z tabeli 2 łatwo stwierdzić, że test T_β zachowuje wartości dwóch parametrów pierwszego testu: czułość diagnostyczna wynosi również 81%, oraz współczynnik chorobowości stanowiący wartość prawdopodobieństwa wstępnego wynosi 39%. Wyższa natomiast jest, w porównaniu z testem T_α , wartość swoistości diagnostycznej, czyli większa jest zdolność testu T_β do potwierdzenia hipotezy H^+ . Wartość tego parametru wynosi:

$$\text{Sw.d.} = \frac{n(PU)}{n(PU + FD)} = \frac{171}{171 + 18} = 90\%$$

Wartość współczynnika predykcji dodatniej wynosi:

$$\text{Pred.}(T_\beta^+) = \frac{n(PD)}{n(PD + FD)} = \frac{98}{98 + 18} = 84,4\%$$

Obliczona wartość współczynnika predykcji będzie jednak różniła się od wartości uzyskanej z wykorzystaniem twierdzenia Bayesa. Otóż zastosowanie tego twierdzenia w projektowaniu kolejnych kroków strategii diagnostycznej umożliwia uwzględnienie prawdopodobieństwa obliczonego we wcześniejszych etapach. Innymi słowy, podstawiając dane do twierdzenia Bayesa dla testu T_β nie uwzględnimy wartości $P(H^+)$ określonej jako współczynnik chorobowości (39%), lecz wartości prawdopodobieństwa końcowego, obliczonego w poprzednim etapie, czyli $P(H^+ / T_\alpha^+) = 62\%$. To byłoby właśnie prawdopodobieństwo, jakie lekarz przypisywałby hipotezie H^+ po uzyskaniu pozytywnego wyniku pierwszego testu w pierwszym etapie diagnozy. Zatem:

$$P(H^+ / T_\beta^+) = \frac{P(H^+ / T_\alpha^+) \times \text{Cz.d.}}{P(H^+ / T_\alpha^+) \times \text{Cz.d.} + (1 - P(H^+ / T_\alpha^+)) \times (1 - \text{Sw.d.})}$$

$$P(H^+ / T_\beta^+) = \frac{0,62 \times 0,81}{0,62 \times 0,81 + (1 - 0,62) \times (1 - 0,90)} = 92,9\%$$

Prawdopodobieństwo końcowe policzone na podstawie wartości predykcji testu T_β wyniosło 84%, natomiast obliczone na podstawie twierdzenia Bayesa 93%. Zastosowanie twierdzenia Bayesa umożliwiło zatem przekroczenie założonego progu potwierdzenia hipotezy H^+ (90%), skutkując decyzją o podjęciu leczenia. Próg ten nie został natomiast przekroczony w projekcie strategii diagnostycznej określającej wartość współczynnika predykcji.

Jak zatem rozumiemy prawdopodobieństwo $P(H^+)$? Personalistycznie? Raczej nie. W pierwszej fazie projektowania strategii wartość ta jest obliczona jako współczynnik chorobowości. Wyliczenie ma na celu ustalenie wartości względnej częstości zdarzenia sprzyjającego w zbiorze wszystkich zdarzeń, tj. ustalamy względną częstość występowania choroby w zbiorze osób poddanych testowi. Nie ma podstaw, aby twierdzić, że jest to prawdopodobieństwo rozumiane personalistycznie. Każda osoba posiadająca dostęp do danych charakteryzujących test może powtórzyć obliczenie uzyskując ten sam wynik. W dalszej kolejności podstawiamy uzyskaną wartość do twierdzenia Bayesa, obliczając zmianę prawdopodobieństwa, która nastąpi na podstawie testu T_α . Widzieliśmy jednak, że zamiast prowadzić obliczenia za pomocą twierdzenia Bayesa możemy wyliczyć wartość predykcji. Wynik będzie ten sam. Obliczając wartość predykcji określamy względną częstość wyniku dodatniego w zbiorze wszystkich wyników dodatnich. Dokonujemy tego na podstawie charakterystyki testu, czyli danych doświadczenia. Ponownie nie mamy podstaw, aby twierdzić, że określamy wartość prawdopodobieństwa w sposób subiektywny. Uzyskana wartość zostaje wykorzystana przy kolejnym zastosowaniu twierdzenia Bayesa, dla potrzeb oceny testu T_β , itd.

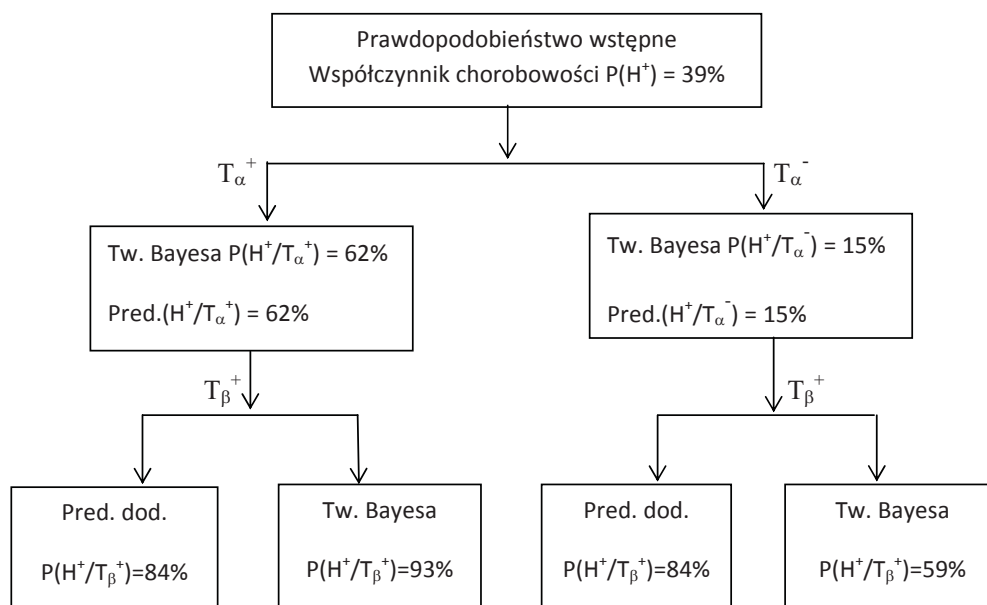
Subiektywny wymiar twierdzenia Bayesa w diagnostyce medycznej nie jest zatem konsekwencją personalistycznego przypisywania wartości prawdopodobieństw a priori. Obliczając wartości współczynnika predykcji w kolejnych etapach testu uwzględniamy stałą wartość prawdopodobieństwa a posteriori występowania choroby w populacji (39%). Natomiast posługując się twierdzeniem Bayesa uwzględniamy w kolejnych etapach testu zmianę prawdopodobieństwa występowania choroby u konkretnego pacjenta, która nastąpiła w efekcie wykonywania testów diagnostycznych na wcześniejszych etapach. Zachowany zostaje jednak główny warunek charakteryzujący prawdopodobieństwo a posteriori – intersubiektywna sprawdzalność. Każdy podmiot, śledząc wcześniejsze etapy diagnozy może sprawdzić, jak zmieniała się wartość prawdopodobieństwa pod wpływem wyników kolejnych testów. Różnica pomiędzy współczynnikiem predykcji a twierdzeniem Bayesa dotyczy w istocie zakresu doświadczeń, które są brane pod uwagę przy obliczaniu prawdopodobieństwa końcowego hipotez. Twierdzenie Bayesa pozwala zrelatywizować końcową wartość prawdopodobieństwa w następujących po sobie testach poprzez uwzględnienie wcześniej uzyskanych wyników. Natomiast w przypadku obliczania współczynnika predykcji w kolejnych etapach procesu diagnozy wpływ wyników wcześniejszych testów na wartość prawdopodobieństwa końcowego nie jest uwzględniany. Można powiedzieć, że w przypadku współczynnika predykcji nie jest zachowana ciągłość zmian wartości prawdopodobieństwa. Każda kolejna wartość predykcji zakłada ten sam punkt wyjścia: ustaloną na stałe wartość prawdopodobieństwa wstępnego, określoną jako częstość występowania choroby w populacji. Mamy zatem do czynienia z dwoma różnymi modelami statystycznymi. Czy jest jednak sens rozpisywać się o różnicach pomiędzy obu metodami, w sytuacji, w której uzyskane wartości różnią się w tak niewielkim stopniu w podanym przykładzie, tj. 84% vs. 93%?

Przeprowadzona analiza dotyczyła projektu dwuetapowej strategii diagnostycznej, w której diagnosta zakładał uzyskanie dwóch dodatnich wyników w następujących po sobie testach T_α i T_β . Projekt ten jest oczywiście niepełny ponieważ nie bierze się pod uwagę, że któryś z tych testów dostarczyłby wyniku ujemnego. Rozważmy zatem

krótko jak zmieniłyby się prawdopodobieństwa określone z wykorzystaniem wartości predykcji i twierdzenia Bayesa wówczas, gdy w T_α uzyskano by wynik ujemny, a następnie w T_β dodatni.

Przyjmując, że prawdopodobieństwo wstępne wynosiło 39%, to ujemny wynik testu T_α zmniejszyłby je do 15,3% (na podstawie twierdzenia Bayesa lub współczynnika predykcji)²⁴. Wówczas dodatni wynik testu T_β , obliczony z wykorzystaniem twierdzenia Bayesa wykazałby 59,4%, podczas gdy wartość predykcji tego testu byłaby taka sama jak w pierwszej strategii diagnostycznej i wynosiłaby 84,4%. Widzimy zatem, że w przypadku zróżnicowania wyników kolejnych testów wybór metody określania strategii diagnostycznej w istotny sposób wpływa na wartości końcowe prawdopodobieństwa, mając kluczowe znaczenie dla podejmowania decyzji medycznych.

Podsumowując dotychczasowe ustalenia, zaprezentowany projekt strategii diagnostycznych, różnicujący twierdzenie Bayesa od wartości predykcji dla testów T_α i T_β miałby następującą postać.



Udział twierdzenia Bayesa i współczynnika predykcji w ocenie prawdopodobieństwa końcowego w projekcie strategii obejmującej dwa testy diagnostyczne: T_α i T_β . Wyraźnie widać, że różnice wyników dotyczą dopiero drugiego etapu diagnostyki, wówczas gdy, posługując się twierdzeniem Bayesa, przyjmujemy obliczone w pierwszym etapie prawdopodobieństwa końcowe jako prawdopodobieństwa wstępne. Różnice te zwiększają się w przypadku uzyskiwania w kolejnych testach odmiennych wyników (dodatnich lub ujemnych).

Schemat 5

²⁴ Podstawą przeprowadzenia obliczeń jest odpowiednio przekształcone twierdzenie Bayesa dla ujemnego wyniku

$$\text{testu } T_\alpha: P(H^+ / T_\alpha^-) = \frac{P(H^+) \times P(T_\alpha^- / H^+)}{P(H^+) \times P(T_\alpha^- / H^+) + P(H^-) \times P(T_\alpha^- / H^-)} = \frac{P(H^+) \times (1 - Cz.d)}{P(H^+) \times (1 - Cz.d) + P(H^-) \times Sw.d.}$$

Schemat 5 pozwala sformułować następującą tezę. Otóż to nie wybór wstępnej wartości prawdopodobieństwa ma charakter personalistyczny, lecz wybór metody rozstrzygającej o końcowej wartości prawdopodobieństwa hipotez diagnostycznych. Jeżeli mianowicie w rozważanym przykładzie diagnosta optymistycznie będzie zakładał, że w sposób trafny wstępnie rozpoznaje chorobę D i formułuje hipotezę H^+ , to może posłużyć się strategią bayesowską, która już w drugim teście pozwoli mu przekroczyć próg leczenia, przy wartości 93%. Przyjmując bardziej ostrożną postawę, być może uznałby, że korzystniejsze jest obliczenie współczynnika predykcji, który w przypadku ujemnego wyniku pierwszego testu (T_{α}^-) pozwoli, mimo tego, w drugim teście (T_{β}^+) uzyskać stosunkowo wysoką wartość prawdopodobieństwa hipotezy H^+ (84%). Dawałoby to szansę na przekroczenie progu leczenia w trzecim etapie diagnozy.

Tezę o tym, że to właśnie wybór pomiędzy strategiami diagnostycznymi ma charakter personalistyczny można dodatkowo uzasadnić w następujący sposób. Otóż diagnosta nie może być optymistą w odniesieniu do możliwości uzyskania pozytywnego wyniku testu T_{α} w sytuacji, gdy $P(H^+)$ wynosi zaledwie 39%, podczas gdy $P(H)$ aż 61%, chyba, że dokonywałby personalistycznej oceny prawdopodobieństwa. Prawdopodobieństwo częstościowe określałoby bowiem niską wartość H^+ , a jednocześnie diagnosta wybierając strategię bayesowską przyjmowałby, że jest duża szansa na uzyskanie pozytywnego wyniku w teście T_{α} - i w efekcie na przekroczenie progu leczenia już w drugim teście. Ale taka sytuacja miałaby małe szanse realizacji w świetle wartości prawdopodobieństwa częstościowego hipotezy H^+ zaledwie na poziomie 39%. Decyzja o wyborze strategii bayesowskiej wyrażałaby w tej sytuacji subiektywne przekonanie diagnosty, nie znajdujące uzasadnienia w wartości prawdopodobieństwa wstępnego, ustalonej w sposób częstościowy.

10. Praktyka stosowania twierdzenia Bayesa w procesie wspomaganie decyzji diagnostycznych

Dokonana analiza może budzić pewne zastrzeżenia. Jaka jest bowiem w rzeczywistości przydatność tego rodzaju ustaleń dla praktyki medycznej? Być może bylibyśmy skłonni zgodzić się, że przyjmujący nas lekarz w pamięci oblicza wartość współczynnika predykcji. Trudno jednak sobie wyobrazić lekarza, który przeprowadzałby obliczenia projektując wieloetapową strategię bayesowską i dodatkowo porównywał ją jeszcze ze strategią opartą na współczynniku predykcji. Taka sytuacja być może ma miejsce w dużych ośrodkach klinicznych, specjalizujących się w określonych rodzajach chorób²⁵. Stosowanie strategii bayesowskiej wydaje się być jednak poza zasięgiem lekarzy pierwszego kontaktu oraz specjalistów przyjmujących w mniejszych ośrodkach, w których większe zróżnicowanie dolegliwości pacjentów uniemożliwia opracowanie jednolitego schematu postępowania diagnostycznego. Zastrzeżenia te nie są jednak trafne. Wbrew pozorom nawet lekarze pierwszego kontaktu dysponują stosunkowo prostą metodą określania wartości prawdopodobieństwa końcowego w analizie bayesowskiej.

²⁵ Np. oddziały chest-pain w USA, w których selekcja pacjentów jest określona właśnie metodą opartą na strategii bayesowskiej – Diamond, Forrester (1979).

Poszukując informacji charakteryzujących efektywność testów diagnostycznych lekarze korzystają z publikacji zamieszczanych w elektronicznych bazach danych. W artykułach tam publikowanych główne parametry testu, czyli czułość i swoistość diagnostyczna, są zestawione ze sobą i prezentowane jako osobny parametr, określany mianem ilorazu wiarygodności (LR)²⁶. Wyróżnia się dodatni i ujemny iloraz wiarygodności. Parametry te precyzowane są w następujący sposób.

$$LR^+ = \frac{Cz.d}{1 - Sw.d}$$

$$LR^- = \frac{1 - Cz.d}{Sw.d}$$

Widać, że iloraz wiarygodności określa stosunek pomiędzy czułością i swoistością diagnostyczną. Łatwo stwierdzić, że w omawianym przykładzie dla dodatniego wyniku testu T_{α} (tabela 1) wartość ilorazu wiarygodności wyniesie:

$$LR(T_{\alpha}^+) = \frac{Cz.d}{1 - Sw.d} = \frac{0,81}{1 - 0,68} = 2,53$$

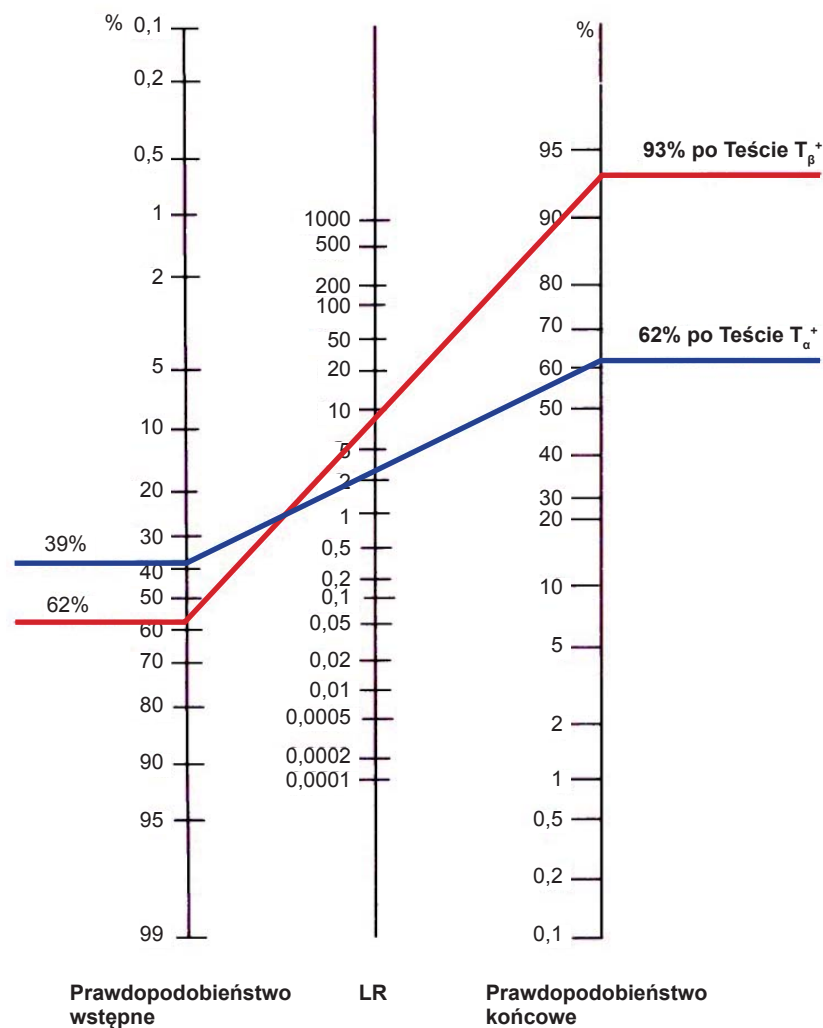
Diagnosta znając z publikacji wartość LR dla danego testu oraz znając wstępną wartość prawdopodobieństwa występowania choroby w populacji²⁷, np. współczynnik chorobowości, może określić prawdopodobieństwo końcowe hipotezy posługując się tzw. nomogramem Fagana²⁸. Nomogram Fagana jest rodzajem wykresu obejmującego trzy kolumny: wartości prawdopodobieństwa wstępnego, wartości LR oraz wartości prawdopodobieństwa końcowego. Łącząc linią prostą wartość prawdopodobieństwa wstępnego z wartością LR, określić można wartość końcową prawdopodobieństwa po teście i następnie użyć jej jako wartości prawdopodobieństwa wstępnego w kolejnym teście. Dla omówionej dwuetapowej strategii diagnostycznej obejmującej wyniki T_{α}^+ oraz T_{β}^+ wykres na nomogramie będzie miał następującą postać²⁹.

²⁶ Halkin, Reichman, Schwaber et al. (1998); Dujardin, van den Ende, van Gompel et al. (1994).

²⁷ Ścisłej, aby w określeniu prawdopodobieństwa końcowego wykorzystać wartość LR, należy wartość prawdopodobieństwa wstępnego wyrazić jako szansę – Jaeschke, Cook, Guyatt (1998): 159. Nie wprowadzam obliczeń, by nie komplikować wyводу. Celem jest wyłącznie ukazanie nomogramu Fagana.

²⁸ Fagan (1975).

²⁹ Schemat za: Fagan (1975); Jaeschke, Cook, Guyatt (1998).



Niebieska linia łączy wstępną wartość prawdopodobieństwa hipotezy H^{+} (39%) z wartością LR^{+} (2,53) dla testu T_{α}^{+} . Wartość końcowa prawdopodobieństwa po teście T_{α}^{+} = 62%. Uzyskany wynik przyjmowany jest następnie jako początkowa wartość prawdopodobieństwa dla testu T_{β}^{+} , dla którego LR^{+} = 8,1 (czerwona linia). Wartość końcowa prawdopodobieństwa po teście T_{β}^{+} wynosi 93%³⁰.

Schemat 6

W analogiczny sposób można na nomogramie zaprezentować – obliczoną wcześniej – strategię z testem ujemnym T_{α} oraz następującym po nim testem dodatnim T_{β} . Nomogram stanowi zatem proste narzędzie projektowania wieloetapowych strategii diagnostycznych³¹. Warto podkreślić, że nomogram może być wykorzystany nie tylko jako reprezentacja obliczeń wykonywanych za pomocą twierdzenia Bayesa, ale również dla potrzeb ustalania wartości predykcyj. W tym przypadku jednak dla danej hipotezy

³⁰ Schemat ma wyłącznie charakter ogólnej ilustracji i nie może stanowić podstawy dla przeprowadzania w procesie diagnozy precyzyjnych ustaleń wartości prawdopodobieństwa hipotez.

³¹ Warto podkreślić, że w ostatnich latach wprowadzone zostały kolejne uproszczenia. Opracowano nomogram określający wartości LR na podstawie czułości i swoistości diagnostycznej, dla sytuacji, w których wartość LR nie jest podana w publikacji – Caraguel, Vanderstichel (2013): 127.

wartość prawdopodobieństwa wstępnego za każdym razem będzie taka sama. W omawianym przykładzie będzie wynosiła 39%.

11. Podsumowanie

W artykule porównane zostały dwie metody projektowania strategii diagnostycznych w medycynie: metoda wykorzystująca twierdzenie Bayesa oraz metoda obliczania współczynnika predykcji. Pokazane zostało, że różnice pomiędzy obu metodami ujawniają się dopiero w projektowaniu wieloetapowych strategii diagnostycznych. Wybór metody projektowania strategii diagnostycznej jest w istocie wyborem pomiędzy dwiema różnymi perspektywami poznawczymi. W pierwszej, zakładanej w twierdzeniu Bayesa, uwzględnia się zmianę prawdopodobieństwa hipotez następującą w efekcie wcześniejszych testów diagnostycznych. W drugiej, zakładanej przy obliczaniu współczynnika predykcji, wcześniejsze zmiany prawdopodobieństwa hipotez nie są brane pod uwagę. Przeprowadzona analiza pozwala stwierdzić, że wybór metody projektowania strategii diagnostycznej i tym samym wybór perspektywy poznawczej ma charakter personalistyczny i w istotny sposób wpływa na podejmowane w procesie diagnozy decyzje.

Bibliografia

- Ajdukiewicz K. (1975), *Logika pragmatyczna*, PWN, Warszawa.
- Caraguel C.B.G., Vanderstichel R. (2013), *The two step Fagan's nomogram: ad hoc interpretation of a diagnostic test result without calculation*, „Evidence Based Medicine” 18 (4): 125–128.
- Diamond G.A., Forrester J.S. (1979), *Analysis of probability as an aid in the clinical diagnosis of coronary artery disease*, „New England Journal of Medicine” 300 (24): 1350–1358.
- Dujardin B., van den Ende J., van Gompel A., Unger J.P., van der Stuyft P (1994), *Likelihood ratios: a real improvement for clinical decision making*, „European Journal of Epidemiology” 10 (1): 29–36.
- Enøe C., Georgiadis M.P., Johnson W.O. (2000), *Estimation of sensitivity and specificity of diagnostic tests and disease prevalence when the true disease state is unknown*, „Preventive Veterinary Medicine” 45 (1–2): 61–81.
- Fagan T.J. (1975), *Nomogram for Bayes's theorem*, „New England Journal of Medicine” 293 (5): 257.
- Fine T.L. (1973), *Theories of Probability. An Examination of Foundation*, Academic Press, New York, London.
- Gigerenzer G., Hoffrage U. (1995), *How to Improve Bayesian Reasoning Without Instruction: Frequency Formats*, „Psychological Review” 102 (4): 684–704.
- Gillies D. (2000), *Philosophical Theories of Probability*, Routledge, London.
- Goodman S.N. (1999), *Toward Evidence-Based Medical Statistics. 1: The P-Value Fallacy*, „Annals of Internal Medicine” 130 (12): 995–1004.
- Halkin A., Reichman J., Schwaber M., Paltiel O., Brezis M. (1998), *Likelihood ratios: getting diagnostic testing into perspective*, „QJM: An International Journal of Medicine” 91 (4): 247–258.
- Howson C., Urbach P. (2006), *Scientific Reasoning. The Bayesian Approach*, Open Court Publishing Company, Chicago, La Salle.
- Hui S.L., Zhou X.H. (1998), *Evaluation of diagnostic tests without gold standards*, „Statistical Methods in Medical Research” 7 (4): 354–370.

- Jaeschke R., Cook D.J., Guyatt G.H. (1998), *Evidence based medicine (EBM), czyli praktyka medyczna oparta na wiarygodnych i aktualnych publikacjach (POWAP)*, „Medycyna Praktyczna” 10 (92): 156–159.
- Jędrychowski W. (1999), *Epidemiologia. Wprowadzenie i metody badań*, Wydawnictwo Lekarskie PZWL, Warszawa.
- Jędrychowski W. (2002), *Podstawy epidemiologii*, Wydawnictwo Uniwersytetu Jagiellońskiego, Kraków.
- Joseph L., Gyorkos T.W., Coupal L. (1995), *Bayesian estimation of disease prevalence and the parameters of diagnostic tests in the absence of gold standard*, „American Journal of Epidemiology” 141 (3): 263–272.
- Moons K.G., van Es G.A., Deckers W., Habbema J.D., Grobbee D.E. (1997), *Limitations of sensitivity, specificity, likelihood ratio, and bayes' theorem in assesing diagnostic probabilities: a clinical example*, „Epidemiology” 8 (1): 12–17.
- Naskalski J. (2005), *Pojęcie normy, wartości referencyjnych i ich znaczenie dla formułowania diagnozy*, [w:] *Diagnostyka laboratoryjna z elementami biochemii klinicznej*, A. Dembińska-Kieć, J. Naskalski (red.), Urban i Partner, Wrocław: 47–57.
- Paulo S., Mendes S., Vizinho R., Carneiro A.V. (2004), *Diagnostic testing, pre- and post-test probabilities, and their use in clinical practice*, „Revista Portuguesa de Cardiologia” 23 (9): 1187–1198.
- Rzepiński T. (2016), *Konstruowanie typologii i klasyfikacji wielowymiarowych dla potrzeb diagnostyki medycznej*, „Zagadnienia Naukoznawstwa” 2 (208): 199–218.
- Sak J. (2013), *Wielowymiarowość postrzegania choroby w kontekście przekonań zdrowotnych i poczucia sensu życia*, Perfecta Info, Lublin.
- Steurer J. (2009), *Pre-test and post-test probabilities. Who cares?*, „Therapeutische Umschau” 66 (1): 5–7.
- Winkler R.L., Smith J.E. (2004), *On uncertainty in medical testing*, „Medical Decision Making” 24 (6): 654–658
- Wulff H.R., Götzsche P.C. (2005), *Racjonalna diagnoza i leczenie. Wprowadzenie do medycyny wiarygodnej czyli Evidence-Based Medicine*, AKTIS, Łódź.