

Andrzej Zieliński

BŁĄD KLASYFIKACJI W BADANIACH EPIDEMIOLOGICZNYCH

Zakład Epidemiologii Narodowego Instytutu Zdrowia Publicznego - Państwowego
Zakładu Higieny

W artykule omówiono błędy popełniane w badaniach epidemiologicznych przy kwalifikowaniu badanych osobników do kategorii narażonych i nie narażonych oraz przy rozpoznawaniu chorób. Podjęto próbę wyjaśnienia wpływu, jaki błędy zaklasyfikowania wywierają na wartości stosunku ryzyka i stosunku szans. Ponadto przedstawiono rolę błędów klasyfikacji narażenia w analizie stratyfikacyjnej mającej na celu wyeliminowanie efektu czynnika zakłócającego.

Słowa kluczowe: błąd pomiaru (klasyfikacji), zmienna kategoryczna, czynnik zakłócający
Key words: measurement error (misclassification), categorical variable, confounder

WSTĘP

Prowadząc badania epidemiologiczne nigdy nie możemy całkowicie ustrzec się popełniania błędów. Mogą one mieć swe źródła w niedoskonałości narzędzi diagnostycznych, lub w czynnikach subiektywnych: braku odpowiedniej staranności, a nawet uprzedzeń czy przesądów badacza, często nieuświadomionych.

Założmy sytuację, w której mamy określoną populację i do niej będziemy chcieli odnieść wyniki naszego badania epidemiologicznego. Mamy też określone kryteria doboru próby do badań. Zawsze, ilekroć na podstawie przyjętych kryteriów próbujemy zaklasyfikować daną osobę lub inny przedmiot badania do określonej grupy musimy liczyć się z możliwością popełnienia błędu.

Jeżeli popełniane przez nas błędy mają charakter przypadkowy, rozkładają się „na obie strony” i przy odpowiednio dużej liczbie pomiarów, średnia pomiarów nie różni się znacznie od średniej rzeczywistej. Znacznie poważniejszy problem stanowią błędy systematyczne, które prowadzą do stronniczości wyników. Takimi właśnie błędami zajmujemy się w tym opracowaniu. Celem jego jest wskazanie na możliwe źródła i skutki błędów związanych z niewłaściwym zaklasyfikowaniem ludzi lub innych przedmiotów do grup badanych, a także z błędnym rozpoznaniem u nich cech pojawiających się w trakcie badania (np. zachorowań). Ten rodzaj błędów znany jako błędy zaklasyfikowania i pomiaru stanowią częste źródło fałszywych wniosków wysnuwanych z prac epidemiologicznych.

Świadomość źródeł błędów i wagi ich następstw jest najważniejszą drogą do ich eliminacji, a przynajmniej ograniczenia ich skutków.

BŁĄD KLASYFIKACJI A BŁĄD POMIARU

Jeśli zmienne, którymi posługujemy się w badaniu epidemiologicznym są zmiennymi kategorycznymi, w wyniku błędu klasyfikacji badana osoba trafią do innej grupy niż powinna. Chory trafią do grupy zdrowych, a narażony do grupy nienarażonych lub na odwrót. Nasze kryteria diagnostyczne możemy przyjąć za narzędzia pomiaru i uznać, że błędne zaklasyfikowanie stanowi swego rodzaju błąd pomiaru. Nie zawsze jednak błąd klasyfikacji polega na niewłaściwym zastosowaniu kryteriów rozpoznania. Bywa, że inne kryteria rozpoznania choroby lub narażenia są stosowane w różnych grupach tego samego badania lub, że zmieniają się w czasie badania. Badający może np. nie dopilnować stosowania tej samej definicji choroby w stosunku do wszystkich uczestników badania. Mamy wtedy do czynienia z błędami klasyfikacji, ale trudno je nazwać błędami pomiaru.

Jeżeli natomiast mamy do czynienia ze zmiennymi ciągłymi np. z poziomem glukozy we krwi, jako potencjalnym czynnikiem narażenia na wystąpienie powikłań cukrzycy, błąd pomiaru może polegać na zawyżeniu lub zaniżeniu wartości liczbowych w obrębie tej samej zmiennej. Jeżeli nie przyjmujemy w naszym badaniu wartości progowych, które dzielą badane osoby na grupy, taki błąd pomiaru nie jest błędem klasyfikacji i wymaga osobnego omówienia.

Podsumowując powyższe uwagi zauważmy, że błąd zaklasyfikowania może wystąpić, gdy niewłaściwie zastosujemy kryteria klasyfikacji popełniając błąd pomiaru lub, gdy zmieniamy kryteria klasyfikacji w czasie badania albo też stosujemy różne kryteria do różnych grup w badaniu. Przy badaniu zmiennych ciągłych podzielonych na kategorie według określonej wartości — progu odciążenia, można popełniać błędy pomiaru, ale nie muszą one prowadzić do błędu klasyfikacji, o ile nie przekraczają one progu. Jeżeli uznamy za próg cukrzycy wartość cukru w surowicy przekraczającą na czczo 124 mg%, to błędne rozpoznanie poziomu cukru 90 mg% jako 110 mg% nie spowoduje błędu klasyfikacji, gdyż zarówno wartość prawidłowa, jak i błędna pozostają po tej samej stronie punktu odciążenia, w tym przypadku w kategorii „zdrowych”.

Błędy klasyfikacji mogą być skutkiem błędów, jakie są popełniane przy stosowaniu testów diagnostycznych. U osoby chorej można nie rozpoznać choroby, a osobą zdrową lub chorą na jakąś inną chorobę można uznać za chorą na chorobę będącą przedmiotem badania epidemiologicznego. Podobna sytuacja występuje, gdy określa się osoby objęte badaniem jako narażone lub nienarażone na analizowane przez nas czynniki. Jeżeli w obrębie reprezentatywnej próby popełnimy błąd w zaklasyfikowaniu chorych i zdrowych albo narażonych i nienarażonych, określony przez nas rozkład badanych cech nie będzie odpowiadał ich rozkładowi w populacji, a zatem badanie nasze będzie miało mniejszą wartość zewnętrzną (*external validity*), czyli mniej zasadne będzie uogólnienie na populację wyników uzyskanych w badaniu próby. Ponadto, należałoby oczekiwać, że ci chorzy, którzy zostali przez nas zaklasyfikowani jako zdrowi, będą mieli częstość występowania czynników narażenia taką jak inni chorzy. Analogicznie będzie się przedstawiała sytuacja z narażeniem zdrowych, fałszywie zakwalifikowanych jako chorzy. W konsekwencji badanie zależności

między narażeniem i chorobą może prowadzić do błędnych wyników. Mamy wtedy do czynienia z obniżeniem wartości wewnętrznej badania (*internal validity*).

Błędy zaklasyfikowania stają się bardziej złożone w miarę zwiększania liczby zmiennych oraz kategorii, które w tych zmiennych występują. Np. w sytuacji najprostszej z możliwych, pobieramy z populacji osoby do badania i klasyfikujemy je jako chore lub zdrowe, czyli mamy do czynienia z jedną zmienną, która dzieli się na dwie kategorie.

	Chorzy - D	Zdrowi - $\approx D$
Zaklasyfikowani jako chorzy: D^*	a	b
Zaklasyfikowani jako zdrowi: $\approx D^*$	c	d
	$a+c$	$b+d$

Czułość (*sen*) rozpoznania stanowi stosunek rozpoznań prawdziwie dodatnich do liczby wszystkich osób chorych. Swoistość (*spec*) rozpoznania stanowi stosunek rozpoznań prawdziwie ujemnych do liczby osób zdrowych.

$$Sen = \frac{a}{a+c} = \frac{a}{D} \qquad Spec = \frac{d}{b+d} = \frac{d}{\approx D}$$

Błędy zaklasyfikowania stanowią wyniki fałszywie ujemne (błąd czułości) „c” oraz wyniki fałszywie dodatnie (błąd swoistości) „b”. Zatem osoby zaklasyfikowane w naszym badaniu jako chorzy, to ludzie naprawdę chorzy tak zaklasyfikowani „a” (*wyniki prawdziwie dodatnie*), i ludzie zdrowi zaklasyfikowani jako chorzy „b” (*wyniki fałszywie dodatnie*). Zaś osoby zaklasyfikowane jako zdrowe, to osoby naprawdę zdrowe, tak zaklasyfikowane „d” (*wyniki prawdziwie ujemne*) oraz osoby chore zaklasyfikowane jako zdrowe „c” (*wyniki fałszywie ujemne*). Tak więc liczba osób zaklasyfikowanych jako chore w naszym badaniu może być wyliczona z następującego równania:

$$D^* = D \text{ sen} + \approx D (1-\text{spec})$$

Odpowiednia formuła dla liczby osób zaklasyfikowanych jako zdrowe:

$$\approx D^* = \approx D \text{ spec} + D (1-\text{sen})$$

Sytuacja, w której zarówno czułość jak i swoistość pomiaru mają dokładnie po 100% zakłada klasyfikację bez żadnych błędów. Jest to ideał osiągany wyjątkowo i to tylko w bardzo prostych pomiarach na niewielkich grupach, najczęściej w stosunku do banalnie prostych problemów. Tak można czasem podzielić dzieci w klasie na chłopców i dziewczynki, ale już nie na dobrych i złych uczniów (1,2).

BŁĄD KLASYFIKACJI A WIARYGODNOŚĆ DANYCH

Problem błędów popełnianych w badaniach epidemiologicznych możemy rozpatrywać w pierwotnym znaczeniu tego słowa jako odchylenie od prawdy, którą w tym przypadku

reprezentuje obiektywna rzeczywistość. Kłopot w tym, że ta obiektywna rzeczywistość nie jest nam dostępna inaczej niż przez nasze i innych badaczy pomiary. Zazwyczaj więc, wynik jednego pomiaru odnosimy do wyników drugiego. W tej sytuacji ważne jest, aby móc ocenić i kwantyfikować zgodność wyników dwóch różnych zaklasyfikowań danych tej samej próby. Jeżeli jedno z tych zaklasyfikowań (pomiarów) ma bardzo wysoką wiarygodność, możemy potraktować je jak złoty standard i do niego odnieść drugi pomiar jak do obiektywnej rzeczywistości. Zazwyczaj jednak mamy do czynienia z dwoma pomiarami o nie do końca ustalonej wiarygodności. Wtedy w pierwszym rzędzie interesuje nas zgodność tych pomiarów ze sobą.

Porównanie takie może nam dostarczyć ważnych informacji na temat jakości narzędzi np. ankiet stosowanych w badaniach analitycznych. Jeżeli np. odpowiedzi na te same pytania ankiety, dwukrotnie zastosowanej w tej samej grupie ludzi, w krótkim odstępie czasu, różnią się znacznie od siebie, powstaje problem, czy ankieta ta została prawidłowo skonstruowana i czy osoby odpowiadające na ankietę są w stanie zrozumieć jej pytania.

Dla zmiennych kategorycznych w ocenie zgodności i wiarygodności wyników odpowiednim wskaźnikiem jest współczynnik *kappa*. W swym najprostszym wydaniu dla dwóch kategorii został on wprowadzony w 1960 r. (3). Później jednak został rozszerzony na zmienne o wielu kategoriach. W zaawansowanych podręcznikach epidemiologii można znaleźć wzory na wyliczenie ważonego współczynnika *kappa* dla wielu kategorii oraz sposoby jego statystycznych oszacowań (4).

Współczynnik *kappa* opiera się na porównaniu zgodności zaklasyfikowania w przypadku dwóch porównywanych pomiarów ze zgodnością, jaką pomiary te osiągnęłyby przy całkowicie losowym rozmieszczeniu elementów w kategoriach, czyli wtedy, gdyby pomiędzy tymi pomiarami nie było żadnej zależności. Prawdopodobieństwo zgodności czysto losowej, czyli przewidywanej, dwu klasyfikacji zmiennych dwuwartościowych podaje następująca formuła:

$$\text{Prawdopodobieństwo przewidywanej zgodności losowej} = p_1 \times p_2 + (1-p_1) \times (1-p_2)$$

gdzie „ p_1 ” oznacza proporcję elementów zaklasyfikowanych jako posiadające daną cechę (np. chorych) w pierwszej niedoskonałej klasyfikacji, a p_2 oznacza proporcję elementów zaklasyfikowanych jako posiadające daną cechę (np. chorych) w drugiej niedoskonałej klasyfikacji. Wtedy iloczyn $p_1 \times p_2$ oznacza losowe prawdopodobieństwo zgodnego zaklasyfikowania „chorych” przez obie klasyfikacje, a iloczyn $(1-p_1) \times (1-p_2)$ zgodnego zaklasyfikowania „zdrowych”.

Zgodność obserwowaną stanowi suma osób zaklasyfikowanych jako chorzy według obu klasyfikacji i zaklasyfikowanych jako zdrowi, również według obu klasyfikacji, podzielona przez całkowitą liczbę obserwacji.

$$\text{Kappa} = (\text{zgodność obserwowana} - \text{zgodność oczekiwana}) / (1 - \text{zgodność oczekiwana})$$

Przykład praktycznego wyliczenia współczynnika *kappa*

Druga klasyfikacja	Pierwsza klasyfikacja		Sumy wierszy
	Chorzy	Zdrowi	
Chorzy	35	18	53
Zdrowi	21	163	184
Sumy kolumn	56	181	N=237

$$p_1 = \frac{56}{237} = 0,236; \quad p_2 = \frac{53}{237} = 0,224;$$

$$\text{Zgodność oczekiwana} = \frac{(56 \times 53) + (181 \times 184)}{237^2} = 0,539$$

$$\text{Zgodność obserwowana} = \frac{35 + 163}{237} = 0,835$$

$$kappa = \frac{0,835 - 0,593}{1 - 0,593} = 0,595$$

Wartość współczynnika *kappa* osiąga "1" przy całkowitej zgodności obu klasyfikacji. Wszyscy osobnicy (lub elementy), którzy posiadają cechę będącą podstawą klasyfikacji według jednego pomiaru, posiadają ją również według drugiego, i wszyscy ci, którzy jej nie posiadają według jednego, nie posiadają jej według drugiego. Wtedy zgodność obserwowana wynosi "1". Natomiast, gdy zgodność obserwowana zaklasyfikowań jest dokładnie równa zgodności oczekiwanej, czyli losowej, wartość współczynnika *kappa* wynosi „0”.

Najprostsza praktyczna interpretacja współczynnika *kappa* jest przyjęta dość arbitralnie:

Wysoka zgodność - $kappa > 0,75$

Akceptowalna zgodność - $0,4 < kappa < 0,75$

Niska zgodność - $kappa < 0,4$

Wartość interpretacyjna współczynnika *kappa* jest wysoka, gdy obie klasyfikacje opierają się na kryteriach wzajemnie od siebie niezależnych. Natomiast tam, gdzie występuje zależność kryteriów obu klasyfikacji, może zdarzyć się tak, iż obie klasyfikacje są obciążone podobnym systematycznym błędem stronniczości pomiaru, a między sobą wykazują wysoką zgodność. Np. Częstość palenia papierosów uzyskana z ankietowania młodzieży, która wypiera się palenia może być zgodna z wynikami ankietowania rodziców, przed którymi palenie jest ukrywane, a oba te wyniki mogą być obciążone wysokim błędem stronniczości pomiaru.

Zatem w interpretacji tego wskaźnika, większą wagę przywiązujemy do wyników niskich, wskazujących na słabą zgodność porównywanych klasyfikacji, niż do wyników wysokich, wskazujących na dużą zgodność. Wysokie wartości *kappa* są na tyle wiarygodne, na ile uda nam się wykluczyć wzajemną zależność kryteriów obu klasyfikacji.

ANALIZA DWÓCH LUB WIĘCEJ ZMIENNYCH KATEGORYCZNYCH

Jeżeli w badaniu są dwie zmienne, z których jedna charakteryzuje np. chorobę, a druga narażenie na określone warunki środowiska, możemy mieć do czynienia z każdą z następujących sytuacji:

- błąd klasyfikacji dotyczy tylko choroby i jest taki sam w grupie narażonych i nienarażonych (błąd nieróżnicowy)
- błąd klasyfikacji dotyczy tylko narażenia i jest taki sam w grupie chorych i zdrowych (błąd nie różnicowy)
- błąd klasyfikacji dotyczy tylko choroby i jest różny w grupie narażonych i nienarażonych (błąd różnicowy)
- błąd klasyfikacji dotyczy tylko narażenia i jest różny w grupie chorych i zdrowych (błąd różnicowy)
- błąd klasyfikacji dotyczy zarówno narażenia i choroby i w obrębie każdej z tych zmiennych może być różnicowy lub nieróżnicowy.

Błąd zaklasyfikowania choroby nazywamy różnicowym, jeżeli występują różnice w wielkości tego błędu (czułości lub swoistości zaklasyfikowania) między kategoriami narażenia. Np. chorobę u narażonych rozpoznajemy z czułością 90%, a u nienarażonych z czułością 80%. Analogicznie błąd zaklasyfikowania narażenia jest różnicowy, gdy w grupie chorych inna jest czułość lub/i swoistość rozpoznań narażenia niż w grupie zdrowych.

Błędy klasyfikacji odbijają się niekorzystnie na miarach porównawczych ryzyka: na stosunku szans, na stosunku ryzyka i na różnicy ryzyka.

Błąd nieróżnicowy powoduje, że wartość liczbowa tego wskaźnika ryzyka względnego oraz stosunek szans zbliżają się do jedności. Wskaźniki powyżej jedności maleją, a poniżej jedności rosną, ale wskaźnik nigdy nie może ulec odwróceniu. Większy od jedności nie stanie się od niej mniejszy, a mniejszy od jedności większym niż jedność. W literaturze anglojęzycznej takie zmiany noszą nazwę “stłumienia w kierunku zera” (*attenuation toward null*). Przy czym zero nie oznacza tu liczby, a hipotezę zerową, która odzwierciedla *status quo*. W kategoriach liczbowych jest to zbliżenie zarówno stosunku szans jak i stosunku ryzyka do jedności. Zatem tego rodzaju błąd nie sprawi, że stosunek szans zostanie odwrócony i czynnik ryzyka zostanie uznany za czynnik ochronny przed chorobą, ale może sprawić, że obniży się moc testu i wynik znamiennej w rzeczywistości nie uzyska w błędnie sklasyfikowanej próbie poziomu znamienności (5).

Wyjaśnienie opisanego wyżej efektu nieróżnicowego błędu klasyfikacji sprowadza się do homogeniczności tabeli 2x2. Tabela doskonale homogeniczna składa się wyłącznie z wartości “oczekiwanych”. Ma ona taki sam stosunek liczb występujących zarówno w kolumnach jak i w rzędach. To znaczy, że stosunek liczby narażonych do liczby nienarażonych jest taki sam wśród chorych, jak wśród zdrowych, a stosunek liczby chorych do liczby zdrowych jest taki sam wśród narażonych jak wśród nienarażonych. Iloczyn szans lub iloczyn ryzyka obliczony na podstawie takiej tabeli da w wyniku wartość 1. Obowiązuje tu gólna zasada: im bardziej proporcje liczb w kolumnach i wierszach tabeli są do siebie zbliżone, tym bliższe jedności są ich iloczyny. Nieróżnicowy błąd klasyfikacji powoduje, że następuje przesunięcie liczb z jednej kolumny do drugiej z zachowaniem proporcji występującej w kolumnie “dawcy”. A zatem proporcja w kolumnie “biorcy” staje się bliższa proporcji występującej u “dawcy” i w ten sposób zwiększa się homogeniczność tabeli. Analogicznie przedstawia się sytuacja,

gdy przesunięcia następują z jednego wiersza tabeli do drugiego z zachowaniem proporcji. Np. jeżeli mamy do czynienia z nieróżnicowym błędem zaklasyfikowania narażenia, taki sam odsetek narażonych chorych jak narażonych zdrowych może być przesunięty do odpowiednich komórek w wierszu (lub kolumnie) nienarażonych.

Jeżeli błąd nieróżnicowy dotyczy dwu lub więcej zmiennych jednocześnie, oczekujemy, że zmiany miar zależności między zmiennymi następują w kierunku ich osłabienia. W przypadku stosunku szans lub stosunku ryzyka wartości tych wskaźników zbliżają się do jedności. Często jednak obniżenie to jest większe, gdy błąd dotyczy tylko jednej zmiennej, niż kiedy dotyczy dwóch zmiennych jednocześnie. Zasadniczą sprawą dla wielkości i kierunku zmian miar epidemiologicznych, przy błędach występujących w więcej niż jednej zmiennej, jest wzajemna niezależność tych błędów.

Gdy błędy popełniane przy pomiarach obu zmiennych są od siebie zależne, mogą się wzajemnie kompensować. Typowym przykładem zależności błędów zaklasyfikowania jest sytuacja, w której badacz zbiera wywiady dotyczące chorób ze stygmatyzacją społeczną. Błąd wynikający z zatajania choroby, może współwystępować z błędem wynikającym z zatajania czynników narażenia na zakażenie, jeżeli są one również przedmiotem stygmatyzacji. Jednak dopóki błąd jest nieróżnicowy, prowadzi on do osłabienia zależności między zmiennymi, ale nie prowadzi do odwrócenia stosunków.

Błąd różnicowy polega na tym, że błędne zaklasyfikowanie choroby lub narażenia jest inne w każdej z kategorii. Innymi słowy czułość lub specyficzność rozpoznania jednej zmiennej jest zależna od wartości innej zmiennej. W takich przypadkach homogeniczność tabeli nie musi się zwiększać i w związku z tym kierunek i zakres zmian, jakie występują w miarach asocjacji, nie jest możliwy do przewidzenia. Np. u narażonych dokładniej rozpoznajemy chorobę niż u nienarażonych i w związku z tym rzadziej popełniamy błędy. Błędy różnicowe mogą prowadzić nie tylko do zaniżania wyników, ale i do ich zawyżania, a w skrajnych przypadkach nawet do odwrócenia stosunku ryzyka. Zaburzenia wywołane przez różnicowe błędy zaklasyfikowania są znacznie groźniejsze dla jakości analizy danych epidemiologicznych, dlatego w planowaniu badań należy zwracać szczególną uwagę na możliwe źródła błędów różnicowych. Dość częstą ich przyczyną jest niejednakowo staranne wykonywanie pomiarów w grupie badanej i w grupie kontrolnej.

Jeżeli jedna zmienna ma więcej niż dwie kategorie, badani osobnicy trafiają już nie tylko do jednego z dwu dychotomicznych przedziałów, ale w obrębie np. przedziału narażonych mogą trafiać do różnych kategorii narażenia. W każdym z tych zaklasyfikowań można popełnić błąd. Gdy kategorie są uporządkowane i błąd zaklasyfikowania polega na przesunięciu do sąsiedniej, najbliższej kategorii i dotyczy takiego samego odsetka badanych osobników we wszystkich kategoriach drugiej zmiennej, będzie to błąd nieróżnicowy. Ale, gdy w takim badaniu próbujemy zmieniać kategorie, np. łączyć je ze sobą, nieróżnicowy błąd zaklasyfikowania może zmienić się na różnicowy. Przy wielu kategoriach zmiennej narażenia, nieróżnicowy błąd, pod warunkiem, że wartość średnia mierzonego narażenia zawsze rośnie ilekroć rośnie narażenie rzeczywiste, prowadzi do przesunięcia stosunku ryzyka i stosunku szans w kierunku jedności. Jednak, jeśli warunek ten nie jest spełniony, może zdarzyć się, że nieróżnicowy błąd klasyfikacji prowadzi do odsunięcia mierzonego stosunku szans od jedności (6,7,8).

BŁĘDY POMIARU W BADANIACH ANALITYCZNYCH

Błąd pomiaru może mieć wpływ na wyniki badania związku cech w analizach epidemiologicznych. W badaniu korelacji zmiennych ciągłych występowanie nieróżnicowych i niezależnych błędów pomiaru zmniejsza wartość współczynnika korelacji, poza wyjątkową sytuacją, gdy korelacja zmiennych jest doskonała i jej współczynnik wynosi 1.

Badania kohort i badania referencyjne (kliniczno-kontrolne) różnią się między innymi tym, w jakiej kolejności określamy w nich narażenie i chorobę. Pierwszy z wymienionych typów badań rozpoczynamy od dobrania osób zdrowych do kohort narażonych i nienarażonych. Błąd, jaki możemy popełnić przy kwalifikowaniu badanych osób do każdej z tych kohort, jest niezależny od zachorowania, które występuje dopiero po włączeniu danej osoby do odpowiedniej kohorty. Dlatego nie ma podstaw, aby zakładać, iż może to być błąd różnicowy.

Jeżeli w badaniach kohort mamy doskonałą (stuprocentową) swoistość, stosunek ryzyka nie ulega zmianie przy obniżeniu czułości. Lecz jeżeli swoistość jest obniżona, stosunek ryzyka ulega przesunięciu w kierunku jedności.

Jednak błędy popełniane w badaniach kohort przy rozpoznawaniu chorób mogą być zależne od wcześniejszego zakwalifikowania badanej osoby jako narażonej lub nienarażonej i w związku z tym mogą mieć charakter różnicowy. Dość typową przyczyną różnicowych błędów zaklasyfikowania chorób w badaniach kohort jest mniejsza dokładność badań diagnostycznych (czułość lub specyficzność) w grupie nienarażonych niż narażonych.

Badania referencyjne (*kliniczno-kontrolne*) rozpoczyna się od rozpoznania przypadków zachorowań i dobrania do nich grupy kontrolnej pochodzącej z tej samej populacji, co przypadki, a zatem składającej się z osób, które, gdyby zachorowały, trafiłyby do grupy przypadków. Wobec tego, że narażenie jest oceniane po rozpoznaniu przypadków, w badaniach referencyjnych istnieje większa możliwość powstania różnicowego błędu klasyfikacji narażenia. Niektórzy badacze mogą mieć nieświadomą skłonność do łatwiejszego rozpoznawania narażenia przypadków niż w grupie kontrolnej, zwłaszcza, jeżeli są już wstępnie przekonani o związku danego narażenia z chorobą, a badanie ma to tylko potwierdzić.

Jeżeli badania referencyjne przeprowadzane są retrospektywnie i dane o narażeniu są uzyskiwane od uczestników badania, może się zdarzyć, że osoby z grupy przypadków lepiej je pamiętają niż zdrowe osoby z grupy kontroli. Osoba chora jest bardziej skoncentrowana na swym zdrowiu, częściej rozważa i lepiej pamięta możliwe przyczyny zachorowania niż osoba zdrowa. Np. lepiej może pamiętać przyjmowanie leków uspokajających w czasie ciąży osoba, której dziecko urodziło się z wadą wrodzoną niż osoba mająca dzieci zdrowe.

W badaniach referencyjnych określenie czułości i specyficzności rozpoznawania narażenia nie wystarcza do określenia stopnia odchylenia stosunku szans w kierunku jedności. Zależy ono również od rozpowszechnienia (*prevalence*) narażenia. Efekt specyficzności silniej zaznacza się przy małym rozpowszechnieniu narażenia, zaś efekt czułości przy dużym. Najmniejszy wpływ ma czułość i specyficzność przy rozpowszechnieniu narażenia w pobliżu 50%, a im dalej od tej wartości, tym efekty te są większe: specyficzności przy wartościach niskich, a czułości przy wysokich (2,9). Wynika to z faktu, iż przy rozpoznawaniu cech rzadko występujących przy danej swoistości badania, liczba wyników fałszywie dodatnich jest wśród wyników dodatnich proporcjonalnie większa. Zaś przy określonej czułości, gdy rozpoznajemy cechy częste, liczba wyników fałszywie ujemnych wzrasta wraz z rzeczywistą częstością występowania badanej cechy.

Jeżeli błąd zaklasyfikowania jest niezależny od drugiej zmiennej i nieróżnicowy, stopień obniżenia stosunku szans może być wyliczony w sposób przybliżony przy zastosowaniu współczynnika $kappa$ (10). Ta część prawdziwego (uzyskanego przy doskonałej selekcji) stosunku szans, która stanowi nadwyżkę ponad 1 pomnożona przez $kappa$ daje nadwyżkę ponad 1 stosunku szans wyliczonego przy błędzie klasyfikacji. Jeżeli $OR=4$, a $kappa=0,75$, to stosunek szans zmniejszony w wyniku błędu klasyfikacji wyniesie: $OR^*=(4-1) \times 0,75 + 1 = 3,25$.

BŁĘDY ZAKLASYFIKOWANIA W OBECNOŚCI CZYNNIKÓW ZAKŁÓCAJĄCYCH

Obecność czynników zakłócających (*confounders*) sama w sobie stanowi wyzwanie dla przeprowadzającego analizę epidemiologiczną. Błąd pomiaru w badaniach, w których występują czynniki zakłócające, może dotyczyć zmiennych analizy (czynników narażenia), zmiennej wynikowej (choroby) oraz samego czynnika zakłócającego, a najczęściej wszystkich trzech wymienionych kategorii zmiennych i w każdej z nich może być różnicowy lub nieróżnicowy.

Czynnik zakłócający, według definicji, stanowi zmienna związana przyczynowo z chorobą i związana statystycznie z narażeniem, ale nie działająca w łańcuchu przyczynowym między narażeniem a chorobą. Innymi słowy czynnik zakłócający nie zmienia siły powiązania między badanym narażeniem a chorobą, ale związek ten zakłóca w inny, opisany wyżej, sposób. Na podstawie tej definicji można wyeliminować efekt czynnika zakłócającego przez stratyfikację uczestników badania w grupach tak, aby w jednej z tych grup czynnik zakłócający występował zarówno u narażonych, jak i u nienarażonych, a w drugiej nie występował zupełnie. Wtedy miary stosunków (stosunek ryzyka i stosunek szans) w obu grupach są jednakowe, dobrze przybliżając miarę rzeczywistą i różnią się od surowej wartości uzyskanej bez stratyfikacji. Jeżeli po dokonaniu stratyfikacji wobec czynnika zakłócającego, nie otrzymamy jednakowych wartości w obu grupach, należy wnosić, że w jednej z tych grup lub w obu, ale w nie jednakowym stopniu, przyjmujemy, że występuje czynnik modyfikujący, zmieniający siłę związku między narażeniem, a chorobą.

Gdy czynnik zakłócający, występujący w badaniu, jest mierzony bez błędu, a pomiar narażenia lub choroby jest obarczony błędem, szczególnie błędem stroniczym wobec kategorii czynnika zakłócającego, może się zdarzyć, że badanie w grupach rozdzielonych według czynnika zakłócającego da różne wyniki w każdej z grup, wskazując fałszywie na występowanie modyfikacji efektu (11).

Daje to złudne wrażenie, że mamy do czynienia z modyfikacją efektu, a nie tylko z czynnikiem zakłócającym. Zatem błąd klasyfikacji może stanowić poważne utrudnienie w analizie mającej na celu wyeliminowanie wpływu czynników zakłócających w badaniach epidemiologicznych.

Popętnienie błędów w zaklasyfikowaniu badanych osobników według zmiennej zakłócającej (*confounder*) sprawia, że stratyfikacja ze względu na tę zmienną rozdziela podzbiory niekompletnie. Jeżeli czynnik zakłócający silnie wiąże się z narażeniem i chorobą, błąd związany z tym faktem może stanowić znaczne odchylenie miar stosunków. Poważne odchylenie stosunku szans może wynikać z jednoczesnego błędu zaklasyfikowania narażenia i czynnika zakłócającego. Może wówczas zaistnieć nawet taka sytuacja, że „surowy”

stosunek szans, mierzony bez stratyfikacji da lepsze przybliżenie stosunku rzeczywistego, niż uzyskany po rozdzieleniu próby według czynnika zakłócającego (2,11).

W większości jednak badań, błędy w przypisaniu czynnika zakłócającego mniej wpływają na wynik ostateczny badania niż te, które dotyczą zaklasyfikowania badanego czynnika narażenia lub choroby. Efekt błędnego zaklasyfikowania zmiennych na odchylenia miar stosunków zależy nie tylko od samej wielkości błędu, ale i od tego jak silny jest związek statystyczny między zmiennymi klasyfikowanymi z błędem. Zasadniczą rolę odgrywa to, czy błędy w zaklasyfikowaniu różnych zmiennych popełniane są niezależnie od siebie, czy są one różnicowe oraz to, jak owe zmienne są rozpowszechnione w populacji.

A Zieliński

CLASSIFICATION BIAS IN EPIDEMIOLOGICAL STUDIES

SUMMARY

The article describes problems caused by measurement bias of categorical variables in epidemiological studies. It analyses attenuation of measures of association, which can be related to misclassification in epidemiological studies. Problems with correction for confounders in the presence of misclassification of exposure are also described.

PIŚMIENNICTWO

1. Willett W. An overview of issues related to the correction of non-differential exposure measurement error in epidemiological studies. *Stat Med* 1989; 8:1031-1040.
2. Rothman KJ, Greenland S. *Modern Epidemiology*. Wyd 2. Philadelphia: Lipincott-Raven Publishers; 1996:125-133.
3. Cohen J. A coefficient for agreement for nominal scales. *Educ Psychol Meas* 1960;20:37-46.
4. Fleiss JL, Cohen J, Everit BS. Large sample standard errors for kappa and weighted kappa. *Psychol Bull* 1969;72:323-7.
5. Armstrong BG. Effect of measurement error on epidemiological studies of environmental and occupational exposures. *Occup Environ Med* 1998;55(10):651-6.
6. Greenland S. The effect of misclassification in the presence of covariates. *Am J Epidemiol* 1988; 112: 564-569.
7. Dosemeci M, Wacholder S, Lubin JH. Does nondifferential misclassification of exposure always bias a true effect toward the null value? *Am J Epidemiol* 1990; 132:746-748.
8. Brenner H. Bias due to non-differential misclassification of polytomous confounders. *J Clin Epidemiol* 1993; 46:57-63.
9. Copeland KT, Checkoway H, Holbrook RH, McMichael AJ. Bias **due** to misclassification in the estimate of relative risk. *Am J Epidemiol* 1977;105:488-495.
10. Thompson WD. Kappa and attenuation of the odds ratio. *Epidemiology* 1990; 1:357-369.
11. Greenland S, Robins JM. Confounding and misclassification. *Am J Epidemiol* 1985a;122:495-506.

Adres autora:

Andrzej Zieliński

Zakład Epidemiologii Narodowy Instytut Zdrowia Publicznego - Państwowy Zakład Higieny

ul. Chocimska 24, 00791 Warszawa

tel 22 5421 204

e-mail: Zielinski@pzh.gov.pl