

Andrzej Zieliński

DOBÓR PRÓBY W BADANIACH EPIDEMIOLOGICZNYCH

Zakład Epidemiologii Państwowego Zakładu Higieny
Kierownik: W. Magdzik

W pracy omówiono problem reprezentatywności próby w badaniach epidemiologicznych. Scharakteryzowano termin reprezentatywność uwzględniając sposoby doboru próby do celu przeprowadzanych badań.

Słowa kluczowe: populacja, próba, reprezentatywność próby, stronniczość selekcji
Key words: population, sample, representativeness of sample, selection bias

POPULACJA I PRÓBA

Fundamentalne w epidemiologii pojęcie reprezentatywności próby nie doczekało się jednolitej definicji i wciąż jest przedmiotem sporów wśród epidemiologów. Przyczyną tego stanu są po części trudności związane z tym, że nawet przy przyjęciu danej definicji, orzeczenie czy dana próba ją spełnia, wymaga dodatkowych badań. Poza tym reprezentatywność próby należy odnosić do typu badania epidemiologicznego w jakim ma być zastosowana. Inne warunki reprezentatywności spełniać winna próba w badaniach przekrojowych, inne w klinicznych badaniach kontrolowanych, a inne w badaniach kliniczno-kontrolnych. Przedmiotem tego opracowania są różne sposoby definiowania *explicite* lub *implicite* reprezentatywności próby oraz wymagania jakie planowanie różnych typów badań stawia doborowi osób do tych badań.

Słownik epidemiologiczny Lasta (1) podaje dwa zasadnicze znaczenia słowa populacja:

1. Wszyscy mieszkańcy danego kraju lub obszaru rozpatrywani łącznie; liczba mieszkańców danego kraju lub obszaru.
2. (Przy dobieraniu próby) Całkowity zbiór jednostek („uniwersum”) z którego próba jest pobrana (wylosowana); nie koniecznie musi to być populacja osób; jednostkami mogą być instytucje, dokumenty lub zdarzenia. Próba ma w założeniu dostarczyć wyników, które są reprezentatywne dla całej populacji.

Bez wielkiej przesady można powiedzieć, że u podstaw epidemiologii jako nauki leżą dwa problemy: pierwszy to problem wiarygodności (*validity*) wyników analiz dokonanych na próbie z populacji, a drugi to problem uogólnienia (*generalizability*) tych wyników na populację, z której pobrana była próba, na inną, podobną populację lub odpowiednio zdefiniowany podzbiór którejś z tych populacji. Wiąże się to z rep-

reprezentatywnością próby. Z terminem tym często mamy do czynienia w życiu politycznym, kiedy ogłaszane są wyniki badań opinii publicznej dokonywane na „reprezentatywnej próbie”, np. tysiąca obywateli. Można przypuszczać, że autorzy badań pod reprezentatywnością próby rozumieją to, że wyniki uzyskane w badanej próbie dają podobny rozkład odpowiedzi, np. poparcia dla różnych partii, jaki by uzyskano, gdyby przebadano wszystkich obywateli Polski powyżej 18 roku życia.

Cel doboru próby reprezentatywnej na ogół nie budzi zastrzeżeń. Jednak o wiele mniejsza zgoda panuje co do sposobu doboru próby reprezentatywnej, a nawet co do samego znaczenia pojęcia reprezentatywności próby. Dwaj wybitni epidemiolodzy i statystycy amerykańscy Kruskal i Mosteller przeprowadzili krytyczną analizę posługiwania się terminem reprezentatywności próby w piśmiennictwie nie naukowym, w piśmiennictwie naukowym z wyłączeniem statystyki oraz w statystyce (2, 3, 4). W publicystyce i w polityce reprezentatywność próby rzadko podawana jest w kontekście, który pozwalałaby na poznanie sposobu doboru próby lub kryteriów na jakich oparta jest jej ocena jako reprezentatywnej. Na ogół „reprezentatywność” stanowi rodzaj znaku jakości i jest używana jako ozdobnik mający podkreślić wiarygodność danych na jakich opiera się publicysta lub polityk. Niekiedy jednak użycie potoczne, poza piśmiennictwem naukowym odwołuje się do sposobów rozumienia reprezentatywności podobnych jak w piśmiennictwie naukowym.

Lista znaczeń i sposoby zastosowań praktycznych określenia „reprezentatywność próby” w piśmiennictwie statystycznym jest bogatsza niż w piśmiennictwie naukowym poza statystyką i oczywiście niż w publicystyce, ale, co ciekawe, nawet statystycy nie są zupełnie wolni od czysto retorycznego posługiwania się tym terminem. Jeśli takie zastosowanie pominąć, Kruskal i Mosteller wymieniają następujące znaczenia terminu reprezentatywność, jakie spotkali w pracach naukowych z dziedziny statystyki:

1. Nieobecność czynników selektywnych w doborze próby
2. Miniatura populacji
3. Przypadki typowe lub idealne
4. Odwzorowania różnorodności populacji
5. Specyficzna metoda dobierania próby
6. Próba pozwalająca na dobre oszacowania badanych parametrów
7. Próba odpowiednia dla celu, któremu ma służyć.

NIEOBECNOŚĆ STRONNICZOŚCI DOBORU PRÓBY

Stronniczość doboru próby (*selection bias*) stanowi systematyczny błąd w badaniach epidemiologicznych groźny dla wiarygodności i możliwości uogólnienia wyników. Polega on na tym, że próba różni się od populacji, z której została wybrana pod względem częstości występowania jakiejś istotnej cechy sprawiającej, że charakterystyka próby ze względu na tę cechę różni się od charakterystyki populacji. Szeroko znanym przykładem stronniczości selekcji było badanie szans wyborczych amerykańskich kandydatów na prezydenta, Rosevelta i Deweya, w roku 1932. Badanie przeprowadzono telefonując do losowo wybranej próby abonentów i pytając na kogo będą głosować. Znaczna większość pytanych opowiedziała się za Deweyem, a wybory wygrał Roosevelt. Losowe wybranie próby nie uchroniło badania przed stronniczością selekcji, gdyż odbywało się ono w obrębie subpopulacji wyselekcjonowanej wstępnie na podstawie posiadania

telefonu. W owym czasie telefony mieli głównie ludzie zamożni, należący do elektoratu Deweya, a nie ubożsi wyborcy Rosevelta, których, jak się okazało, było więcej.

Nieobecność stronniczości stanowi w swej istocie kryterium negatywne i jako takie jest mało skuteczne w zastosowaniach praktycznych. Jeżeli stwierdzimy stronniczość selekcji, możemy z pewnością powiedzieć, że próba nie jest reprezentatywna. Natomiast nie stwierdzenie stronniczości niczego nie przesądza. Stronniczość mogła zostać przeoczona, tym bardziej, że jest wiele rodzajów stronniczości doboru próby, które trudno nieraz uchwycić intuicyjnie.

Aby zwiększyć możliwość nie stronniczego doboru próby stosowane są losowe metody selekcji, czyli takie w których każda jednostka z populacji ma takie samo prawdopodobieństwo trafienia do próby, a prawdopodobieństwa wylosowania poszczególnych jednostek są od siebie niezależne. Czy jest to jednak metoda, która zawsze prowadzi do próby dobrze odtwarzającej populację? Weźmy populację składającą się w połowie z mężczyzn, a w połowie z kobiet. Prawdopodobieństwo wylosowania równych lub bliskich sobie liczb kobiet i mężczyzn można wyliczyć z rozkładu dwumianowego i jak wiadomo zależy ono od liczebności próby. Dla jednej pary wynosi ono zaledwie 50%.

Czasem więc, aby zapewnić bliższe podobieństwo próby i populacji stosowana jest metoda mieszana: próbkowanie kwotowe. Chcąc wylosować próbę w równych proporcjach składającą się z kobiet i mężczyzn możemy wybrać losowo pewną liczbę kobiet wśród kobiet i taką samą liczbę mężczyzn wśród mężczyzn. Metoda taka może zagwarantować odtworzenie w próbie rozkładu niektórych cech w populacji, jednak w stosunku do innych cech nie objętych kwotowaniem możliwa jest przy tej metodzie bardzo duża stronniczość. Jak łatwo przewidzieć im więcej jest kategorii, których proporcje z populacji chcemy uwzględnić w próbie tym próba musi być odpowiednio większa, co podraża koszt badania. Jeśli mamy ograniczone możliwości powiększenia próby, liczebność poszczególnych kategorii staje się zbyt niska dla efektywnej analizy statystycznej.

MINIATURA POPULACJI

Rezultatem wolnego od wszelkiej stronniczości doboru byłaby próba, w której rozkład wszystkich cech odpowiada dokładnie rozkładowi tych samych cech w populacji. Taki ideał próby reprezentatywnej jest jednak trudny do osiągnięcia ze względu zarówno na fluktuacje przypadkowe jak i systematyczne błędy stronniczości. Najbliższa tego ideału byłaby odpowiednio liczna próba losowa, ale i taki dobór próby mógłby być obciążony dużymi odchyleniami, szczególnie w stosunku do rzadkich chorób. Liczebność próby, która oddawałaby prawdopodobieństwo rozkładu choroby pojawiającej się w populacji raz na sto tysięcy musiałaby wynosić kilka milionów.

Kolejny problem stanowi kryterium zgodności rozkładów populacji i próby. Aby orzec tę zgodność musielibyśmy znać rozkłady wszystkich cech populacji, a to wykracza poza cel najambitniejszych nawet badań. Lepsze przybliżenie realnych sytuacji badań epidemiologicznych daje wyważona definicja Stephana i McCarthego (5): „Próba reprezentatywna jest to próba, która dla określonego zbioru zmiennych odwzorowuje populację... [wykonane w jej obrębie] pewne specyficzne analizy... (obliczenia średnich, odchyłeń standardowych itp___) dostarczają wyniki położone w akceptowalnych przedziałach wokół odpowiednich wartości występujących w populacji, z wyjątkiem

rzadkich sytuacji, gdy wyniki te wypadają poza granicami tych przedziałów." W definicji Stephana i McCarthego odwzorowanie dotyczy tylko tych cech, które stanowią przedmiot badania i ma ono charakter przybliżeń statystycznych.

Usiłowanie odwzorowania populacji przez stratyfikację doboru próby według takich zmiennych jak płeć, grupy wieku, obszar geograficzny itp. może co prawda zbliżyć rozkład w próbie do rozkładu w populacji ale, jak to podkreślono wyżej, w miarę uwzględniania kolejnych zmiennych wzrastają wymagania w stosunku do liczebności próby szybko osiągając wielkości mało realne.

PRÓBA REPREZENTATYWNA JAKO TYPOWA

Jest to określenie mało precyzyjne i z kontekstów w jakich bywa stosowane wynikają różne sposoby jego rozumienia. Zwykle idea typowości opiera się na przekonaniu, że można wybrać jakiś fragment większej całości, który posiada niektóre ważne cechy i zależności między nimi takie same jak całość. Na ogół typowość cech próby polega na wysokiej częstości ich występowania w populacji lub na tym, że ich wartości bliskie są średniej dla populacji. W tym ostatnim znaczeniu typową próbę mógłby reprezentować nawet jeden osobnik, ale nie byłaby to próba bardzo użyteczna w badaniach epidemiologicznych. Z tym wiąże się pewne niebezpieczeństwo dobrze znane statystykom. Bardzo różne rozkłady prawdopodobieństwa mogą mieć te same średnie, a bardzo podobne rozkłady, w których występują wyniki skrajne, dalekie od średniej, mogą mieć różne średnie.

Podstawowa trudność w wykorzystaniu takich „typowych” prób w badaniach epidemiologicznych polega na tym, że rzadko tak wybrana próba jest typowa pod każdym względem. Powiat o takim samym jak w danych zbiorczych dla Polski rozkładzie liczby kobiet i mężczyzn, grup wieku oraz ludności wiejskiej i miejskiej może mieć bardzo czyste powietrze i wodę, a może mieć środowisko zanieczyszczone substancjami toksycznymi i wyższy od przeciętnej odsetek palących papierosy. Nie musi mieć zatem typowej dla reszty kraju umieralności. Są jednak badania, w których wybór typowej próby stanowi podstawę dla bardzo wartościowego warsztatu epidemiologicznego uzyskującego wyniki o wielkiej doniosłości. Przykładem takiego badania jest kohorta Framingham. Uniwersalność wyników badania Framingham wynika z tego, że zależności tam badane mają w populacji ludzkiej charakter powszechny. Może inne miasteczko miałoby inną częstość występowania otyłości i nadciśnienia tętniczego, ale zależności pomiędzy otyłością i nadciśnieniem tętniczym, a chorobami serca są dla populacji ludzkiej uniwersalne, a odchylenia od tych zależności stanowią anomalie mogące być przedmiotem odrębnych badań (6). Typowość Framingham stanowi użyteczną podstawę do szacowania zagrożeń zdrowia większych populacji o podobnym przekroju demograficznym, ale uniwersalność wyników tego badania pochodzi ze starannie dobranych grup ekspozycji i grup kontrolnych, które wcale nie są i nie muszą być „typowe” ani reprezentatywne dla jakiegokolwiek większej populacji w znaczeniach wymienionych powyżej.

ODWZOROWANIE RÓŻNORODNOŚCI POPULACJI

Jeżeli populacja jest podzielona na kategorie (a która nie jest?) pokrycie populacji przez próbę rozumiane jest jako włączenie do próby reprezentantów lub przynajmniej

jednego reprezentanta każdej z kategorii. Według Tippetta: „Jest oczywiste, że próba może być reprezentatywna dla populacji tylko jeśli wszystkie części populacji mają szanse do niej trafić” (7). W epidemiologii nie pobieramy próby z urny z kolorowymi kulami. W realnym życiu nie wszystkie jednostki są jednakowo łatwo dostępne, a często już na wstępie są pogrupowane z racji wieku, wspólnoty zamieszkania, grupy zawodowej lub innych cech. Dobrą ilustrację poboru próby ze zróżnicowanej populacji daje Wadley: „[reprezentatywność] jest najlepiej osiągana, gdy pobieramy próbę z tak wielu części populacji jak to tylko możliwe. Jeśli, na przykład, owoce w sadzie mają być ocenione na podstawie próby 1000 jabłek, lepiej jest wziąć po 50 jabłek z każdego z 20 dobrze rozłożonych drzew niż po 250 jabłek z każdego z czterech drzew.” (8) Ten sam autor zwraca uwagę, że dobrym sprawdzianem reprezentatywności doboru prób jest ich powtarzalność czyli niski błąd standardowy próbkowania.

Reprezentatywność rozumiana jako pokrycie różnorodności populacji jest wyraźnie odmienna od miniatury populacji, gdyż nie stawia wymogu proporcjonalnej reprezentacji wszystkich podgrup, a tylko włączenie do grupy ich przedstawicieli. Jest też wyraźnie sprzeczna z rozumieniem reprezentatywności jako doboru przypadków typowych. Nie o typowość lub przeciętność tu chodzi, a raczej o odwzorowanie różnorodności.

Należy podkreślić, że ten rodzaj reprezentatywności lepiej sprawdza się w poprawności politycznej niż w epidemiologii i rzadko bywa stosowany przez statystyków. W dążeniu do uzyskania w próbie przedstawicieli wszystkich grup łatwo można doprowadzić do różnicowej stronniczości selekcji, i w konsekwencji do poważnych błędów w oszacowaniu miar epidemiologicznych.

REPREZENTATYWNOSĆ DEFINIOWANA PRZEZ PODANIE METODY DOBORU PRÓBY

Prezentacja badań epidemiologicznych w formie sprawozdania lub publikacji powinna w rozdziale dotyczącym metod zawierać dokładnie omówiony sposób dobierania jednostek do badań, czyli tworzenie próby. Nazywanie tak scharakteryzowanej próby reprezentatywną nie wnosi do opisu nic nowego. Większość epidemiologów i statystyków stawia wymóg, aby próba była losowo pobrana z całej populacji lub, w analizie stratyfikacyjnej, z poszczególnych jej warstw. W tym ostatnim przypadku liczebność jednostek w poszczególnych warstwach w próbie powinna odpowiadać proporcji ich liczebności w populacji. Jeżeli z braku środków, czy z powodu nierównomiernej dostępności, próba stanowi odejście od wzorca próby losowej, czy warstwowej próby losowej, nie musi to całkowicie przekreślać wyników badań. Podanie metody doboru próby pozwala zorientować się w stopniu zagrożenia błędem selekcji i skłania do ostrożniejszej interpretacji wyników badań. Należy pamiętać, że we współczesnej medycynie zastosowanie gotowych pakietów statystycznych przesunęło główną groźbę popełnienia błędu z analizy danych na dobór próby.

PRÓBA UMOŻLIWIAJĄCA WIARYGODNE OSZACOWANIA PARAMETRÓW POPULACJI

Sukhatme Pandurang tak definiuje reprezentatywność: „Metoda próbkowania, jeżeli ma dostarczyć próby reprezentatywnej dla populacji, włącznie z tym, że zmienność

jednostek w populacji jest odwzorowana w próbie tak dokładnie jak wielkość próby na to pozwala tak, że wiarygodne oszacowania cech populacji mogą być uzyskane z próby" (9). Taka definicja ma jednak cechy błędnego koła. Po to dobieramy próbę, aby oszacować parametry populacji. Zatem trudno opierać reprezentatywność pobieranej próby na tym, co mamy za pomocą tej próby dopiero określić.

PRÓBA REPREZENTATYWNA JAKO WYSTARCZAJĄCA DO ZAŁOŻONEGO CELU

Jest to definicja minimalistyczna, ale trudno odmówić jej rozsądku. Nie zawsze, aby rozwiązać postawiony problem musimy badanie wykonywać na próbie losowej lub na miniaturze populacji. Jeśli w próbie stu kolejnych pacjentów poradni lekarskiej przy fabryce chemicznej znajdziemy dziesięć osób z kontaktowym zapaleniem skóry to mimo, iż nie jest to próba losowa możemy z dużym prawdopodobieństwem orzec, że robotnicy tej fabryki mają styczność z substancją o dużym potencjale uczulającym. Trzeba tu jednak zastrzec się od razu, że badania na próbach dobranych bez wielkiej dyscypliny metodologicznej mogą dać nam odpowiedzi na stosunkowo proste pytania, dotyczące zjawisk częstych lub nowych w danym obszarze. Im bardziej wnikliwie chcemy zbadać problem tym wyższe wymagania musimy postawić doborowi próby.

W badaniach epidemiologicznych stosowane są rozmaite metody doboru próby - próbkowania. Niekiedy są one specyficznie dostosowane do rodzaju badania.

Proste próbkowanie losowe występuje wtedy, gdy każda jednostka w populacji ma jednakowe prawdopodobieństwo być wylosowaną do próby. Znając liczebność populacji „n” i zakładaną liczebność próby „p” z tabeli lub generatora liczb losowych wybiera się p kolejnych liczb z przedziału $\{1 < p < n\}$. Zbiór jednostek, które mogą być wylosowane do próby nosi nazwę przestrzeni próbkowania (*sampling frame*). W założeniu powinna ona być identyczna z populacją, ale nie zawsze tak się dzieje. Jeśli chcemy zebrać opinie mieszkańców jakiegoś miasta i opieramy się na wywiadach zbieranych telefonicznie przestrzeń próbkowania ograniczona jest do posiadaczy telefonów, którzy stanowią tylko część populacji miasta.

O jakości badania w dużym stopniu decyduje też dobór jednostki (*sampling unit*). Jeśli chcemy zbadać rozkład liczby dzieci w rodzinach możemy wylosować kobiety zamężne i pytać je o posiadaną liczbę dzieci lub same dzieci i pytać je o liczbę posiadanego rodzeństwa. W tym drugim przypadku istnieje większa szansa wylosowania dzieci z rodzin wielodzietnych i otrzymania rozkładu odchyłonego w kierunku rodzin z większą liczbą dzieci.

Próbkowanie systematyczne polega na uszeregowaniu jednostek i następnie wybiieraniu ich w równych odstępach. Może to być co dziesiąty pacjent poradni cukrzycowej, lub co trzeci noworodek, który przyszedł na świat w danym szpitalu. Jest to sposób wygodny, łatwy w zastosowaniu i nie wymaga ustalenia przestrzeni próbkowania przed rozpoczęciem selekcji. Jakość próby zależy od wstępnego ułożenia jednostek, z których dobieramy próbę. Jeśli dokonywano by wyboru wśród żołnierzy ustawionych według wzrostu, a próba byłaby mała, to moglibyśmy się spodziewać nadreprezentacji żołnierzy wysokich lub niskich zależnie od tego z którego końca szeregu rozpoczęlibyśmy selekcję. Aby sprawdzić, czy wyniki uzyskane w metodzie systematycznego próbkowania zbliżają się do próby losowej należy dobrać więcej niż jedną próbę i porównać ich

wariancje. Jeśli nie różnią się wiele jest szansa, że nasz sposób próbkowania jest zadowalający dla analizy epidemiologicznej.

Próbkowanie warstwowe (stratyfikacyjne). W tym sposobie doboru próby, populacja jest podzielona na warstwy (strata) czyli grupy jednostek mających pewne wspólne cechy np. przedział wieku, a następnie proste próby losowe pobiera się z poszczególnych warstw. Próbkowanie warstwowe ma wiele zalet w porównaniu z prostym doбором losowym. Po pierwsze badacz ma zapewnioną reprezentację każdej podgrupy populacji, po wtóre populacja jest podzielona na podgrupy o większej homogenności, a zatem mniejszej wariancji niż populacja jako całość, co daje bardziej precyzyjne oszacowania. Próbki pobrane z warstw o większej wariancji mogą być odpowiednio liczniejsze w celu poprawienia precyzji oszacowań. Największym ograniczeniem próbkowania warstwowego jest mała liczebność prób pochodzących z niektórych warstw, jeśli liczba dostępnych jednostek jest ograniczona.

Próbkowanie grupowe (*cluster sampling*) polega na wstępnym wyborze grup zamiast jednostek, a następnie badanie wszystkich jednostek w obrębie grupy. Można wylosować klasy szkolne, jednostki wojskowe czy zakłady pracy, a następnie badać wszystkie osoby z tak wylosowanej grupy. Jeśli na przykład chcemy zbadać rozpowszechnienie w danym mieście próchnicy zębów wśród dzieci szkolnych, łatwiej będzie wylosować kilka klas w szkołach tego miasta i przebadać uczęszczające tam dzieci kompletnie, niż losować próbę ze wszystkich dzieci w tym mieście. Takie badanie jest znacznie tańsze i dla wielu zastosowań daje zupełnie wystarczające oszacowania. Najczęściej przy tej samej liczbie jednostek wariancja jest większa w tak dobranej próbie niż w losowej, ale można to skompensować zwiększeniem liczby analizowanych grup.

Próbkowanie wielostopniowe stanowi odmianę próbkowania grupowego, z tą różnicą, że w tym wypadku w obrębie wybranych grup dokonywane jest próbkowanie losowe, a w próbkowaniu grupowym analizowane są wszystkie jednostki grupy.

Ocena stosunku prób. W niektórych sytuacjach dla oceny wiarygodności metody doboru próby dobrze jest znaleźć kryterium pozwalające na sprawdzenie dokładności danej metody wyszukiwania przypadków. Jeśli próbujemy ocenić liczbę złamań szyjki kości udowej operowanych w danym roku w danym mieście możemy skorzystać z rejestru wypisów ze szpitala. Lecz niektóre wypisy mogą pomijać rozpoznanie przy wypisie, np. u pacjentów przeniesionych do oddziału intensywnej terapii z powodu zatorów płucnych, którzy są wypisywani z rozpoznaniem „zator tętnicy płucnej”. W takim przypadku wyszukanie pewnej liczby (np. 100) przypadków złamania szyjki w książce operacyjnej i następnie sprawdzenie ile z tych osób miało postawione rozpoznanie złamania szyjki przy wypisie pozwoli na oszacowanie odsetka błędów w rozpoznaniach wypisowych i dopiero na tej podstawie wyliczenie przybliżonej liczby osób leczonych z powodu tej choroby w przyjętym okresie czasu.

Niekiedy epidemiolodzy stosują próbkowanie dwuetapowe. Najpierw w badaniu pilotowym dokonywane jest wstępne oszacowanie parametrów i na tej podstawie dobierana jest liczebność próby, a często i sposób próbkowania w badaniu zasadniczym.

SPECYFIKA DOBORU PRÓBY W BADANIACH KLINICZNO-KONTROLNYCH

W badaniach kliniczno-kontrolnych punktem wyjścia jest dobór grupy przypadków. Przypadki mogą być gromadzone w instytucjach medycznych na podstawie stawianych rozpoznań lub w innych okolicznościach, które dają możliwość wybrania osobników posiadających badaną cechę (np. chorobę). Taka próba bez względu na definicję nie jest próbą reprezentatywną dla populacji, jest wyselekcjonowana ze świadomą stronniczością ze względu na pewną określoną cechę lub zespół cech. Jak zatem dobrać w tym badaniu reprezentatywną grupę kontrolną i co w tej sytuacji oznaczać może reprezentatywność grupy? **Głównym celem w doborze kontroli jest, aby rozkład czynnika, (-ów) narażenia w grupie kontrolnej był taki sam jak w populacji, z której pochodzą przypadki.** Można to uzyskać, lub raczej zbliżyć się do założonej sytuacji, dobierając kontrole zgodne z badanymi przypadkami pod względem możliwie wielu cech za wyjątkiem cechy badanej. Czynniki narażenia działają w czasie i przestrzeni dlatego kontrole powinny: pochodzić z tej samej okolicy co przypadki i być dobierane w tym samym, a przynajmniej nie odległym czasie, mieć podobny wiek, status społeczny i tryb życia, a w wielu badaniach również te samą płęć lub rozkład płci.

Podręczniki epidemiologii podają wiele praktycznych wskazówek jak uzyskać reprezentatywność grupy kontrolnej przez unikanie stronniczości selekcji. Jeżeli przypadki znajdowane są w szpitalu, a grupa kontrolna pochodzi z populacji osób nie hospitalizowanych, może wystąpić stronniczość selekcji wynikająca z tego, że dane narażenie cechuje wszystkie osoby hospitalizowane, a nie wyłącznie hospitalizowane z określonego powodu. W takich sytuacjach należy kontrole dobierać spośród innych hospitalizowanych. Doskonałym przykładem doboru grupy kontrolnej w badaniach kliniczno-kontrolnych jest klasyczna praca Dolla i Hilla nad wpływem palenia tytoniu na powstawanie raka płuc. W odpowiedzi na krytykę ze strony ekspertów wynajętych przez firmy tytoniowe, że porównują przypadki hospitalizowane z nie hospitalizowanymi kontrolami, dobrali oni jako grupę kontrolną osoby hospitalizowane z podejrzeniem raka płuc, u których to rozpoznanie wstępne nie zostało potwierdzone (10).

O niektórych badaniach kliniczno-kontrolnych możemy powiedzieć, że są „zagnieżdżone” (*nested*) w kohorcie. Taką kohortą może być grupa zawodowa wybrana ze względu na wykonywany zawód lub miejsce pracy lub inny zespół osób dobrany tak jak jest to robione w badaniach kohort. Wtedy z tej samej kohorty pochodzą przypadki i możemy przyjąć, że rozkład narażenia kontroli jest taki sam jak w populacji, z której pochodzą przypadki ponieważ jest to ta sama populacja (nasza kohorta).

DOBÓR DOPASOWANY (*MATCHING*)

Dopasowywanie w badaniach epidemiologicznych zostało wprowadzone w przekonaniu, że będzie to procedura eliminująca wpływ czynników zakłócających i wzmacniająca wykrywane związki pomiędzy narażeniem i jego następstwem, np. chorobą (11). W dokładniejszej analizie okazało się jednak, że jest to procedura zwiększająca wydajność badań bez wpływu na siłę asocjacji między cechami (12). Może być ono stosowane w badaniach kohort, gdzie w kategoriach dopasowania dobieramy odpowiednią liczbę narażonych i nie narażonych. Jeśli kryterium dopasowania jest, np. płęć, dobieramy w odpowiednim stosunku narażonych mężczyzn i narażone kobiety i w takim samym stosunku mężczyzn i kobiety nie narażone. Potencjalnym źródłem błędów w dopasowa-

nych kohortach może być związek z kryterium dopasowania konkurujących czynników ryzyka lub przyczyn wypadania z badań. Tym niemniej w dopasowanych badaniach kohort zakłócający efekt kryterium dopasowania jest eliminowany. O wiele bardziej ryzykowne jest stosowanie dopasowywania w badaniach kliniczno-kontrolnych. Jeżeli kryterium dopasowywania jest jednocześnie czynnikiem ryzyka lub jest powiązane statystycznie z, innymi niż badany, czynnikami ryzyka, prowadzi to do stronniczości selekcji i najczęściej do zaniżenia wyniku. Wyjściem jest stosowanie analizy warstwowej w kategoriach kryterium dopasowania. Np. jeśli kryterium dopasowania jest płeć należy wykonać osobne analizy dla kobiet i dla mężczyzn. I zawsze należy sprawdzać czy kryterium dopasowania jest statystycznie neutralne zarówno wobec narażenia jak i choroby.

A Zieliński

SAMPLE SELECTION IN EPIDEMIOLOGICAL STUDIES

SUMMARY

Problems with sample selection in epidemiological studies starts with the definition of representativeness of the sample. The definition of representativeness as a "miniature of the population" is impractical, especially for very rare disease, and not suitable for certain types of epidemiological studies. More plausible are operational definitions, which refer to certain statistical features of the sample or to the sampling method suitable for the study design. The article reviews different meanings of representativeness used directly or implicitly in epidemiological and statistical studies and gives examples of certain methods of sample selection for the specific purposes including matching in design. Some important sources of selection bias and methods of avoiding it are also addressed in this article.

PIŚMIENNICTWO

1. Last JM. A dictionary of epidemiology. 3 wyd. New York: Oxford University Press; 1995.
2. Kruskal W, Mosteller F. Representative sampling, I: non-scientific literature. *Internat Statistical Rev* 1979;47:13-24.
3. Kruskal W, Mosteller F. Representative sampling, II: scientific literature, excluding statistics. *Internat Statistical Rev* 1979;47:111-27.
4. Kruskal W, Mosteller F. Representative sampling, III: the current statistical literature. *Internat Statistical Rev* 1979;47:245-65.
5. Stephan FF, McCarthy PJ. Sampling options: an analysis of survey procedure. New York: Wiley, 1958.
6. Dawber TR, Moore FE, Mann GV II. Coronary heart disease in the Framingham study. *Am J Public Health* 1957;47:4-24.
7. Tippett LHC. The methods of statistics. 4 wyd. New York: Dover, 1952.
8. Wadley FM. Experimental statistics in entomology. Washington DC: US Dept Agriculture, 1967.
9. Sukhatme Pandurang V. Sampling theory of surveys with applications. Ames: Iowa State U. Press, 1954.

10. Doll R, Hill AB A study of the aetiology of the carcinoma of the lung. *BMJ* 1952;2:1271-86.
11. Miettinen OS. Matching and design efficiency in retrospective studies. *Am J Epidemiol* 1970;91:111-8.
12. Kupper LL, Karon JM, Kleinbaum DG, i in. Matching in epidemiological studies: validity and efficiency considerations. *Biometrics* 1981;37:271-92.

Adres autora:

Andrzej Zieliński
Zakład Epidemiologii Państwowego Zakładu Higieny
ul. Chocimska 24, 00-791 Warszawa