

Kazimierz Drosik

# Badania kliniczne z losowym doborem chorych — nauka, wiara czy tylko reklama?

Randomized clinical studies — science, belief or advertisement only

**Artykuł jest tłumaczeniem pracy:**

 Drosik K. Randomized clinical studies — science, belief or advertisement only. *Oncol Clin Pract* 2020; 16. DOI: 10.5603/OCP.2020.0022.

Należy cytować wersję pierwotną.

**Adres do korespondencji:**

 Dr n. med. Kazimierz Drosik  
 ul. Lawendowa 28, 45–470 Opole  
 Tel. 506 197 837  
 e-mail: drosik@go2.pl

**STRESZCZENIE**

Wyniki badań klinicznych z losowym doborem chorych są istotne w ocenie skuteczności leków i nowych metod leczenia. W publikacjach w czasopismach medycznych i prezentacjach konferencyjnych duże znaczenie mają analizy statystyczne wyników. W niniejszym artykule przedstawiono informacje dotyczące roli i wartości niektórych elementów analizy statystycznej. Zwrócono uwagę, że w badaniach klinicznych istotni są chorzy i czas do zdarzenia, będący punktem końcowym. W związku z tym powstaje pytanie: czy liczne analizy statystyczne są rzeczywiście potrzebne, czy też stanowią pewien rodzaj promocji leku?

**Słowa kluczowe:** badania kliniczne, analizy statystyczne

**ABSTRACT**

The results of randomized clinical studies are important in evaluation of drugs' or medical procedures' efficacy. Statistical analyses are meaningful in publications and conference presentations. This paper discusses the role and value of selected statistical methods. It is clear that in clinical studies patients and time to event (recognized as end study point) are essential. According to that the question arises — do all these statistical analyses are important or do they play only a role in drug promotion.

**Key words:** clinical studies, statistical analyses

 Copyright © 2020 Via Medica  
 ISSN 2450–1646

*Oncol Prakt Klin Edu* 2020; 6: 187–194

**Wprowadzenie**

Badania kliniczne z losowym doborem chorych odgrywają najważniejszą rolę w ocenie nowych leków i nowych metod leczenia oraz ustalaniu standardów postępowania w medycynie. Choć wyniki tych badań są powszechnie akceptowane, to jednak ich interpretacja może budzić wątpliwości. Główny problem stanowią zastosowane metody statystyczne, które są często mało zrozumiałe dla przeciętnego czytelnika lub niechętnie analizowane wskutek bezgranicznego zaufania wobec autorów i ich interpretacji.

W ocenie wartości nowych leków przeciwnowotworowych i innych metod postępowania wykorzystuje się

badania różnych faz o różnych celach. Celem może być ustalenie maksymalnej dawki tolerowanej, ocena aktywności leku lub określenie skuteczności leku w postaci ustalenia jego wpływu na czas przeżycia chorych.

Choroby nowotworowe to poważny problem kliniczny i społeczny, ponieważ w wielu wypadkach powodują skrócenie spodziewanego czasu życia. Kluczowe znaczenie ma określenie wpływu leku lub zastosowanej metody leczenia na wydłużenie czasu przeżycia całkowitego. Badania kliniczne, w których punktem końcowym jest czas przeżycia, są więc najważniejsze i stanowią ostateczne podsumowanie badań faz wcześniejszych. Pewną alternatywą czasu całkowitego przeżycia jako punktu końcowego jest czas do nawrotu lub czas do progresji choroby.

Chociaż trwa dyskusja dotycząca wyższości jednego punktu końcowego nad drugim, to dla celów niniejszej publikacji nie jest to istotne. Ważne jest, że i w jednym, i w drugim przypadku mierzy się czas od wejścia chorego do badania (w badaniu z doborem losowym powinna to być data randomizacji) do momentu zdarzenia, jakim może być nawrót choroby, progresja lub zgon. Metodyka badania, przynajmniej od strony statystycznej, jest w tych wypadkach taka sama. Różnice zawierają się tylko w możliwości określenia czasu zdarzenia — czas zgonu jest jednorazowym punktem na osi czasu, niezależnym od założeń badania. Nawrót lub progresja choroby również stanowi jednorazowe zdarzenie na osi czasu, najczęściej jednak jest stwierdzana we wcześniej zaplanowanych okresach, w których wykonywane są badania kontrolne. Jeżeli więc wyniki badania będą przedstawione w formie wykresu, to krzywa prezentująca czas do wystąpienia progresji lub nawrotu choroby będzie miała przebieg skokowy, podczas gdy krzywa czasu przeżycia całkowitego będzie raczej ciągła. Z tego względu łatwiej będzie omówić problem interpretacji wyników badania na modelu badania, w którym punktem końcowym jest nawrót lub progresja choroby.

## Badanie z losowym doborem chorych

Uwarstwienie (stratyfikacja) i losowy dobór chorych (randomizacja)

Największą trudność w zaplanowaniu i przeprowadzeniu badania klinicznego, w którym punktem końcowym jest czas do zdarzenia, stanowi to, że nie można ustalić wcześniej, w którym momencie zdarzenie może lub powinno wystąpić. Gdyby było wiadomo, ile czasu przeżyje nieleczony chory do wystąpienia zdarzenia, to łatwo byłoby wykazać, o ile dłuższy będzie czas przeżycia po zastosowaniu badanego leku lub innej metody terapii. Każdy chory uczestniczący w badaniu byłby swoją własną kontrolą. Tak jednak niestety nie jest i dlatego badanie kliniczne opiera się na porównaniu wyników w badanej grupie chorych z wynikami w grupie kontrolnej. Warunek podstawowy stanowi to, aby grupy chorych były tak podobne, żeby jedyną różnicę między ramieniem badanym i kontrolnym stanowił zastosowany lek (skojarzenie leków) lub zastosowana metoda leczenia.

Najlepiej byłoby przeprowadzać badania na bliźniętach jednojajowych, jednak nawet w tym wypadku istniałaby wątpliwość, czy u bliźniąt wszystkie zdarzenia są takie same i występują w tym samym czasie. Wątpliwości narastają, gdy badanie jest przeprowadzane z udziałem osób niespokrewnionych. Aby w obu ramionach badania uczestniczyli chorzy maksymalnie porównywalni, wprowadzono zasadę uwarstwienia (stratyfikacja) i losowego przydziału chorych do poszczególnych ramion (randomizacja). Celem stratyfikacji jest równy rozdział chorych

pod względem czynników rokowniczych o znanym wpływie na wystąpienie zdarzenia stanowiącego punkt końcowy. Czynniki, które powinny być ujęte w stratyfikacji, jest oczywiście coraz więcej, co wynika z postępu wiedzy o biologii danej choroby. Celem randomizacji jest równy rozdział z uwagi na nieznanne czynniki rokownicze. Zakłada się, że poprzez losowy rozdział chorych również te czynniki o nieznanym wpływie na zdarzenie rozłożą się jednakowo w obu ramionach. Wymienione czynniki z definicji są nieznanne, co powoduje brak możliwości ich określenia w danym czasie i uzyskania rzeczywistej porównywalności charakterystyki chorych w obu ramionach badania. Można jedynie w nią wierzyć. Zamiast dowodu istnieje więc tylko wiara, że dowodem dysponujemy. W tym momencie ważne jest zwrócenie uwagi na prawidłowość kwalifikowania chorych do badania. Kwalifikowanie chorych niespełniających w pełni kryteriów włączenia w celu umożliwienia uczestnictwa „za wszelką cenę” może w oczywisty sposób wpłynąć na wynik. Nie powinna również mieć miejsca indywidualizacja podejmowania decyzji o włączaniu chorych do badania — zasadą jest, że w ośrodku uczestniczącym w badaniu każdy chory spełniający kryteria włączenia powinien być kwalifikowany, o ile wyraża na to zgodę. Jednak każdy chory, który spełnia kryteria, ale nie został włączony do badania, niezależnie od przyczyny, może mieć wpływ na końcowy wynik oraz ocenę jakości badania. Oczywiście niedopuszczalne jest prowadzenie dwóch lub więcej badań o identycznych kryteriach kwalifikacji. Przydział chorych do różnych równoległe prowadzonych badań na podstawie decyzji lekarza wypacza całkowicie sens randomizacji. Ten ważny błąd jest niestety trudny do wykrycia, można w tym zakresie jedynie apelować do etyki badacza.

## Przebieg badania

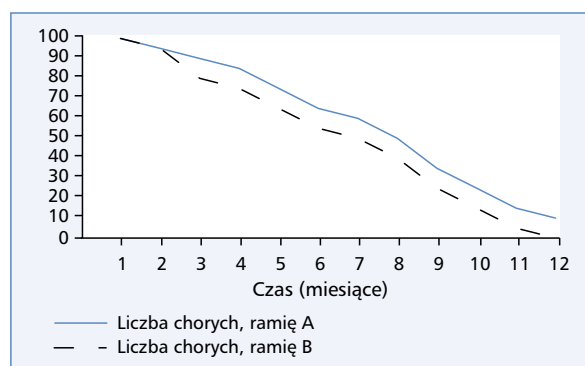
Chory włączony do badania, przydzielony w drodze stratyfikacji i randomizacji do ramienia badanego lub kontrolnego, otrzymuje odpowiednie leczenie — nowy lek (lub skojarzenie) lub nową metodę terapii, a w grupie kontrolnej przykładowo jest to leczenie standardowe. W wypadku badania, którego punkt końcowy stanowi nawrót lub progresja choroby, badania kontrolne wykonywane są w regularnych i z góry ustalonych odstępach czasu. W razie stwierdzenia wystąpienia ocenianego zdarzenia chory kończy udział w badaniu, choć nadal jest poddawany obserwacji pod względem toksyczności. Ten aspekt w obecnym opracowaniu został całkowicie pominięty, ponieważ metody oceny toksyczności są proste i nie wymagają szczególnej wiedzy. Problemem jest interpretacja oceny skuteczności leku, kombinacji leków lub innej metody leczenia.

W badaniu oceniającym skuteczność leku lub innej metody leczenia mamy tylko chorego i czas do zdarzenia. Wszyscy chorzy musieli spełnić kryteria włączenia, w każdym z ramion są więc podobni, jednak w obrębie tego

samego ramienia populacja chorych jest zróżnicowana chociażby pod względem czynników stratyfikujących. Należy o tym pamiętać.

W wyniku kolejnych badań kontrolnych chorzy, u których stwierdzono zdarzenie stanowiące punkt końcowy, są wyłączeni z dalszej oceny. Najlepiej prześledzić to na prostym wykresie, w którym na osi rzędnych mamy liczbę chorych, a na osi odciętych czas (odstęp czasu, w których wykonywane są badania kontrolne). Byłby to najprostszy i najbardziej rzeczywisty sposób przedstawienia wyniku badania. Nie musimy w tym momencie sięgać do przykładowego badania klinicznego, gdyż możemy samodzielnie utworzyć wiele modeli takiego badania i sporządzić wykresy na powyższej zasadzie. Jeżeli jako podstawę przyjmemy wykresy jakiegoś rzeczywistego badania, to możemy również opracować model tego badania. Co prawda w prezentacji wyniku rzeczywistego badania nie ma liczb bezwzględnych, a tylko krzywe prawdopodobieństwa zdarzenia, jednak w tym momencie interesuje nas tylko kształt przebiegu krzywych. Ten problem zostanie wyjaśniony w dalszej części opracowania. Jeśli zamiast prawdopodobieństwa wstawimy faktyczne liczby chorych, to stwierdzimy, że w obu ramionach stopniowo zmniejsza się liczba chorych żyjących nadal bez zdarzenia, tyle że w badaniu „pozytywnym” liczba chorych bez zdarzenia będzie się zmniejszała szybciej w ramieniu kontrolnym. W wielu badaniach te różnice nie są jednak duże. Możemy stworzyć wiele modeli, które potem możemy przyrównać do faktycznie opublikowanych badań. Przykładem może być badanie, w którym liczba początkowa chorych jest jednakowa w obu ramionach. Przy pierwszych — przykładowo dwóch — ocenach

liczba zdarzeń jest jednakowa w obu ramionach. Na wykresie obie krzywe się pokrywają. Założmy teraz, że w trzeciej ocenie okaże się, że liczba zdarzeń jest większa w ramieniu kontrolnym. Krzywe na naszym wykresie się rozsuną (tzw. separacja krzywych) o wielkość, jaka stanowi różnicę liczby chorych, u których stwierdzono dane zdarzenie w ramieniu badanym i kontrolnym. Jeżeli w kolejnych ocenach liczba chorych bez zdarzenia będzie się zmniejszała w obu ramionach o tę samą wielkość, to krzywe będą przebiegać równolegle i będą dalej „odseparowane”. Sprawia to wrażenie, że nadal występują różnice między ramionami, chociaż w rzeczywistości tych różnic nie ma, bo liczba zdarzeń w obu ramionach jest taka sama. W ramieniu kontrolnym jest tylko o tyle mniej chorych bez zdarzenia, o ile było więcej zdarzeń w tym ramieniu przy trzeciej ocenie (tab. 1, ryc. 1). W tym



Rycina 1. Dane zawarte są w tabeli 1

Tabela 1. W kolumnach A1 i B1 przedstawiono liczby chorych rozpoczynających dany okres obserwacji, w kolumnach A2 i B2 — liczby chorych, u których stwierdzono zdarzenie w danym okresie, a w kolumnach A3 i B3 — liczby chorych bez zdarzenia na koniec ocenianego okresu

| Czas (miesiące) | Liczba chorych |    |     |         |    |     |
|-----------------|----------------|----|-----|---------|----|-----|
|                 | Ramie A        |    |     | Ramie B |    |     |
|                 | 1              | 2  | 3   | 1       | 2  | 3   |
| 1               | 100            | 0  | 100 | 100     | 0  | 100 |
| 2               | 100            | 5  | 95  | 100     | 5  | 95  |
| 3               | 95             | 5  | 90  | 95      | 15 | 80  |
| 4               | 90             | 5  | 85  | 80      | 5  | 75  |
| 5               | 85             | 10 | 75  | 75      | 10 | 65  |
| 6               | 75             | 10 | 65  | 65      | 10 | 55  |
| 7               | 65             | 5  | 60  | 55      | 5  | 50  |
| 8               | 60             | 10 | 50  | 50      | 10 | 40  |
| 9               | 50             | 15 | 35  | 40      | 15 | 25  |
| 10              | 35             | 10 | 25  | 25      | 10 | 15  |
| 11              | 25             | 10 | 15  | 15      | 10 | 5   |
| 12              | 15             | 5  | 10  | 5       | 5  | 0   |

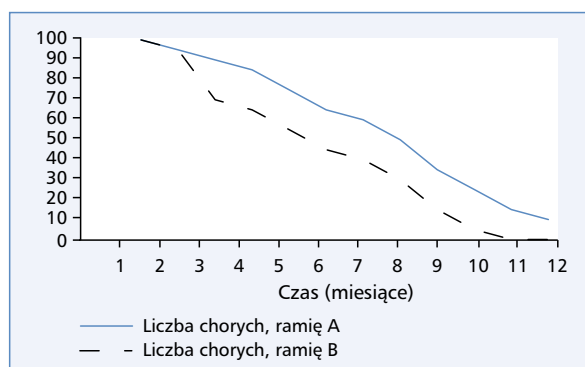
modelu różnica między zdarzeniami w ramionach A i B przy trzeciej ocenie wynosi 10 chorych. Kiedy zdarzenie w ramieniu kontrolnym wystąpi u środkowego chorego, krzywa dla tego ramienia przetnie linię poprowadzoną ze środka rzędnych. Krzywa dla ramienia badanego przetnie tę linię z opóźnieniem. Powiemy, że dla ramienia badanego wydłużyła się mediana czasu do zdarzenia.

Zróbmy następny eksperyment. W ramieniu kontrolnym zwiększymy liczbę chorych, u których przy trzeciej ocenie wystąpiło zdarzenie. Krzywe się bardziej rozsuną i równocześnie zwiększy się różnica między medianami (tab. 2, ryc. 2). Stąd wynika, że zwiększenie lub zmniejszenie różnicy między medianami zależy przede wszystkim od różnic w liczbie zdarzeń w obu ramionach. Jeśli zmienimy moment oceny (nie trzecia ocena, ale którakolwiek kolejna), w którym stwierdzono różnicę, to okaże się, że nie ma to wpływu na różnicę między medianami. W dalszym ciągu ta różnica będzie zależała tylko od różnic w liczbie zdarzeń. Taką sytuację przedstawiono w tabeli 3 i na rycinie 3. Ważne jest oczywiście to, aby te różnice wystąpiły w pierwszej połowie ogólnej liczby chorych uczestniczących w badaniu. Różnice te mogą się ujawniać przy każdej kolejnej ocenie i wówczas wpływ na medianę będzie miała ich suma.

Widzimy więc, że mediana jest raczej miernikiem liczby zdarzeń, a nie czasu, w którym te zdarzenia mają miejsce.

W utworzonym modelu mamy nadal dwie — odseparowane — krzywe, które przebiegają równolegle względem siebie. Zakładamy, że krzywe się rozdzieliły, ponieważ chorzy w ramieniu eksperymentalnym otrzy-

mali bardziej skuteczne leczenie. Co się jednak stanie, gdy skuteczne leczenie zostanie zastosowane jedynie u chorych z ramienia kontrolnego, którzy spowodowali różnicę liczby zdarzeń w trzeciej ocenie (przykład pierwszy) lub w ocenie kolejnej (przykład drugi)? Jeżeli to leczenie jest bardziej skuteczne, to u tych chorych nie powinno wystąpić zdarzenie, a więc krzywa dla ramienia kontrolnego nie przesunie się w dół. Krzywe będą się nadal pokrywały. Powyższa sytuacja uwidacznia twierdzenie, że mała liczba chorych może zadecydować o końcowym wyniku badania. Gdyby końcowy wynik badania był przedstawiony w liczbach bezwzględnych, tak jak na proponowanych modelach, byłaby jasność co do tego, jaka jest faktyczna skuteczność nowego leku, skojarzenia leków lub innej — nowej — metody leczenia. Wynik byłby widoczny na tych prostych wykresach bez zastosowania jakiegokolwiek statystyki.



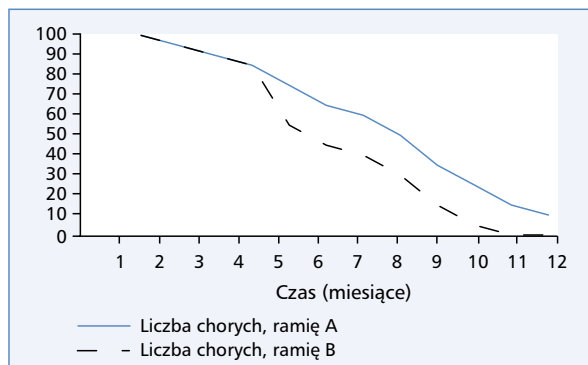
Rycina 2. Dane zawarte są w tabeli 2

Tabela 2. W kolumnach A1 i B1 przedstawiono liczby chorych rozpoczynających dany okres obserwacji, w kolumnach A2 i B2 — liczby chorych, u których stwierdzono zdarzenie w danym okresie, a w kolumnach A3 i B3 — liczby chorych bez zdarzenia na koniec ocenianego okresu

| Czas (miesiące) | Liczba chorych |    |     |         |    |     |
|-----------------|----------------|----|-----|---------|----|-----|
|                 | Ramię A        |    |     | Ramię B |    |     |
|                 | 1              | 2  | 3   | 1       | 2  | 3   |
| 1               | 100            | 0  | 100 | 100     | 0  | 100 |
| 2               | 100            | 5  | 95  | 100     | 5  | 95  |
| 3               | 95             | 5  | 90  | 95      | 25 | 70  |
| 4               | 90             | 5  | 85  | 70      | 5  | 65  |
| 5               | 85             | 10 | 75  | 65      | 10 | 55  |
| 6               | 75             | 10 | 65  | 55      | 10 | 45  |
| 7               | 65             | 5  | 60  | 45      | 5  | 40  |
| 8               | 60             | 10 | 50  | 40      | 10 | 30  |
| 9               | 50             | 15 | 35  | 30      | 15 | 15  |
| 10              | 35             | 10 | 25  | 15      | 10 | 5   |
| 11              | 25             | 10 | 15  | 5       | 5  | 0   |
| 12              | 15             | 5  | 10  | 0       | 0  | 0   |

Tabela 3. W kolumnach A1 i B1 przedstawiono liczby chorych rozpoczynających dany okres obserwacji, w kolumnach A2 i B2 — liczby chorych, u których stwierdzono zdarzenie w danym okresie, a w kolumnach A3 i B3 — liczby chorych bez zdarzenia na koniec ocenianego okresu

| Czas<br>(miesiące) | Liczba chorych |    |     |         |    |     |
|--------------------|----------------|----|-----|---------|----|-----|
|                    | Ramie A        |    |     | Ramie B |    |     |
|                    | 1              | 2  | 3   | 1       | 2  | 3   |
| 1                  | 100            | 0  | 100 | 100     | 0  | 100 |
| 2                  | 100            | 5  | 95  | 100     | 5  | 95  |
| 3                  | 95             | 5  | 90  | 95      | 5  | 90  |
| 4                  | 90             | 5  | 85  | 90      | 5  | 85  |
| 5                  | 85             | 10 | 75  | 85      | 30 | 55  |
| 6                  | 75             | 10 | 65  | 55      | 10 | 45  |
| 7                  | 65             | 5  | 60  | 45      | 5  | 40  |
| 8                  | 60             | 10 | 50  | 40      | 10 | 30  |
| 9                  | 50             | 15 | 35  | 30      | 15 | 15  |
| 10                 | 35             | 10 | 25  | 15      | 10 | 5   |
| 11                 | 25             | 10 | 15  | 5       | 5  | 0   |
| 12                 | 15             | 5  | 10  | 0       | 0  | 0   |



Rycina 3. Dane zawarte są w tabeli 3

Na zakończenie obecnej części warto sprawdzić na modelach liczbę możliwości przebiegu krzywych zdarzeń dla tej samej różnicy median. W każdym przypadku stwierdzimy, że zależy to zawsze od liczby zdarzeń (nie zawsze od czasu, w którym te zdarzenia nastąpiły).

W publikacjach przedstawiających wynik badania klinicznego w postaci wykresów prezentujących czas przeżycia całkowitego lub czas do innego punktu końcowego zamiast liczb bezwzględnych widzimy krzywe prawdopodobieństwa.

### Metoda oszacowania Kaplana–Meiera

Wynik badania przedstawiony w odniesieniu do liczb bezwzględnych byłby łatwo zrozumiały. Zasadniczą wadą

takiego rozwiązania jest jednak to, że taka analiza byłaby możliwa dopiero po zakończeniu udziału w badaniu przez wszystkich chorych (po wystąpieniu u wszystkich chorych zdarzenia będącego punktem końcowym). Takie badanie trwałoby bardzo długo, co dotyczyłoby szczególnie leczenia adiuwantowego. Z tego powodu do analizy badania stosuje się oszacowanie (tzw. estymator) Kaplana–Meiera, czyli rachunek dla obserwacji niedokończonych [1, 2].

Do badania klinicznego włączani są kolejni chorzy, u których ocenia się czas do wystąpienia zdarzenia. Prospektywnie określa się różnicę między ramionami, która będzie miała znaczenie. Na podstawie dotychczasowych danych o liczbie i czasie występowania zdarzeń u chorych leczonych metodą, która zostanie zastosowana u chorych w ramieniu kontrolnym, ustala się liczbę chorych potrzebną do udowodnienia postawionej tezy, że różnica w zdarzeniach osiągnie zakładany poziom i będzie statystycznie znamienne. Zwróćmy uwagę na fakt, że zakłada się, iż zdarzenia będą występowały w obu ramionach, ale w założonym okresie tych zdarzeń będzie mniej w ramieniu badanym. Badanie będzie prowadzone nie do momentu, w którym u wszystkich uczestników wystąpi dane zdarzenie przyjęte jako punkt końcowy (zgon, nawrót lub progresja choroby), ale do uzyskania założonej liczby zdarzeń. Oczywiście chorzy uczestniczący w badaniu, u których nie wystąpi zdarzenie (punkt końcowy), będą kontynuowali leczenie, co jednak nie będzie już przedmiotem zasadniczej analizy.

W czasie, kiedy stwierdzona zostanie założona liczba zdarzeń, czyli badanie jako takie zostanie zakończone, chorzy włączeni do badania będą się cechowali różnym czasem obserwacji. Najdłuższy charakteryzuje tych



włączonych do badania na początku, najkrótszy zaś — włączonych na krótko przed zakończeniem badania. Ze względu na różny czas udziału w badaniu ocena dokonana w odniesieniu do liczb bezwzględnych byłaby niemożliwa. Wykorzystuje się więc szacowanie metodą Kaplana–Meiera, czyli określa się prawdopodobieństwo przeżycia określonego czasu bez zdarzenia — liczby bezwzględne są zastąpione prawdopodobieństwem. Przyjmijmy, że do każdego ramienia włączono 100 chorych. Po każdym miesiącu udziału w badaniu następuje ocena. Jeżeli wszyscy chorzy przeżyli pierwszy i drugi miesiąc bez zdarzenia, to po dwóch miesiącach w każdym ramieniu będzie nadal 100 osób. Prawdopodobieństwo przeżycia bez zdarzenia zostanie obliczone w ten sposób, że liczbę chorych, którzy przeżyli dany okres bez zdarzenia, dzielimy przez liczbę chorych, którzy dany okres rozpoczęli. W tym wypadku będzie to  $100 : 100 = 1$  dla pierwszego i drugiego miesiąca. Gdyby jednak w trzeciej ocenie (po 3 miesiącach) u 5 chorych w ramieniu badanym stwierdzono oceniane zdarzenie, czyli bez zdarzenia przeżyłoby 95 chorych, to prawdopodobieństwo przeżycia danego okresu bez zdarzenia wyniesie  $95 : 100 = 0,95$ . Jeśli w tym samym czasie w ramieniu kontrolnym stwierdzi się 15 zdarzeń, prawdopodobieństwo przeżycia bez zdarzenia w ramieniu kontrolnym wyniesie  $85 : 100 = 0,85$ . Prawdopodobieństwo przeżycia bez zdarzenia obliczane jest dla każdego okresu oddzielnie. Do następnego okresu oceny wejdzie 95 chorych w ramieniu badanym i 85 chorych w ramieniu kontrolnym. Jeśli w następnej ocenie stwierdzi się zdarzenie u np. 4 chorych w ramieniu badanym, to prawdopodobieństwo przeżycia bez zdarzenia tego okresu wyniesie  $91 : 95 = 0,96$ . Aby jednak przeżyć te 4 kolejne miesiące, chory musiał przeżyć pierwsze 3 miesiące. Z tego powodu mnoży się przez siebie prawdopodobieństwo przeżycia kolejnych miesięcy. Prawdopodobieństwo przeżycia 4 miesięcy wyniesie więc  $1 \times 1 \times 0,95 \times 0,96 = 0,91$ . Jeśli z kolei w ramieniu kontrolnym w czwartym miesiącu stwierdzonych zostanie np. 6 zdarzeń, to prawdopodobieństwo przeżycia czwartego miesiąca wyniesie  $79 : 85 = 0,93$ , a prawdopodobieństwo przeżycia 4 miesięcy wyniesie  $1 \times 1 \times 0,85 \times 0,93 = 0,79$ . W ten sposób, mnożąc przez siebie prawdopodobieństwa przeżycia kolejnych okresów, otrzymujemy prawdopodobieństwo przeżycia dla całego czasu obserwacji. Zwróćmy jednak uwagę na fakt, że w momencie zakończenia rzeczywistego badania liczba chorych ocenianych w poszczególnych okresach będzie malała nie tylko z powodu ubywania chorych z powodu wystąpienia zdarzenia, ale również z tego względu, że chorzy później włączeni do badania cechują się niewystarczającym czasem obserwacji. Przykładowo, chory, którzy uczestniczy w badaniu dopiero od 6 miesięcy, nie może być brany pod uwagę w obliczeniu prawdopodobieństwa przeżycia 8 miesięcy i okresu dłuższego. Na tym polega przewaga oszacowania metodą Kaplana–Meiera nad analizą opartą na liczbach

bezwzględnych. Na wykresie krzywe prezentujące wynik badania w liczbach bezwzględnych są zastąpione krzywymi prawdopodobieństwa przeżycia. Gdyby badanie zakończyć dopiero po wystąpieniu zdarzenia u ostatniego chorego i wynik przedstawić w postaci krzywych opartych na liczbach bezwzględnych, to te krzywe powinny się pokryć z otrzymanymi dużo wcześniej krzywymi prawdopodobieństwa. W publikacjach badań z randomizacją pod wykresem prezentującym krzywe prawdopodobieństwa przeżycia znajduje się — aczkolwiek nie zawsze — tabela pokazująca bezwzględne liczby chorych, które były podstawą obliczenia prawdopodobieństwa przeżycia dla danego okresu. Na tej podstawie możemy zobaczyć, że chociaż krzywa pokazuje prawdopodobieństwo przeżycia 20% chorych, to obliczono to na podstawie przeżycia czasem tylko 1 albo 2 chorych. Z przebiegu krzywych można jednak odczytać to samo, co znajdowałoby się na wykresach liczb bezwzględnych, czyli możliwe jest obliczenie różnic w zdarzeniach w poszczególnych okresach między ramionami. Różnicę między medianami możemy dostrzec w tym momencie, w którym prawdopodobieństwo przeżycia w każdym ramieniu wyniesie 0,5. Nie zmienia to faktu, że różnica ta nadal będzie zależała przede wszystkim od różnicy liczby zdarzeń pomiędzy ramionami. Z tego powodu często spotykane twierdzenie, że zastosowane „nowe” leczenie wydłuża czas przeżycia o różnicę pomiędzy medianami, nie znajduje uzasadnienia. Być może rzeczywiście u jakiegoś chorego o tyle wydłużył się czas przeżycia do zdarzenia, ale na pewno nie dotyczy to wszystkich chorych. Aby można było stwierdzić, że nowa metoda leczenia wydłuża czas przeżycia wszystkich chorych o wielkość mediany w badaniu, u chorych z obu ramion zdarzenie nie występowałoby przez jakiś czas, następnie musiałyby wystąpić w tym samym momencie u wszystkich chorych z ramienia kontrolnego i u żadnego z ramienia badanego, po czym po jakimś czasie, odpowiadającym różnicy median, zdarzenie wystąpiłoby równocześnie u wszystkich chorych z ramienia badanego. Realność takiej sytuacji trudno sobie nawet wyobrazić.

W większości — nawet „pozytywnych” — badań większość chorych cechuje się takim samym czasem przeżycia w obu ramionach w poszczególnych okresach oceny. Ostateczna różnica stwierdzona w badaniu jest sumą różnic w kolejnych ocenach. Przypuszczenie, że wszyscy chorzy z ramienia badanego odnieśli korzyść, nie znajduje uzasadnienia. Tego rodzaju stwierdzenia, które jednak często spotykamy w prezentacjach, są jedynie zwykłą reklamą leku lub metody.

## Znamiennosc statystyczna w badaniach klinicznych

W celu uwiarygodnienia wyników badania stosuje się testy znamiennosci (istotności) statystycznej. Nawet jeśli różnica jest statystycznie znamienna, nie znaczy to

jeszcze, że jest istotna klinicznie. Z taką sytuacją mamy do czynienia, gdy rzeczywista różnica jest niewielka.

Jeżeli w każdym z dwu ramion badania byłoby 10 chorych i w wyniku zastosowania nowego leku w ramieniu badanym zdarzenie wystąpiłoby tylko u 1 chorego wobec 9 w ramieniu kontrolnym, to fakt istnienia różnicy byłby widoczny „gołym okiem” i zbędne byłyby testy statystyczne. Jeśli jednak w jednym ramieniu wystąpiłoby 5 zdarzeń, a w drugim 6, to byłyby wątpliwości, czy ta różnica nie jest przypadkowa. W takim wypadku niezbędne są testy statystyczne i oczywiście także zwiększenie liczby chorych koniecznej do udowodnienia różnicy. To uwzględnia się już na poziomie planowania badania. Zakładając wielkość istotnej klinicznie różnicy, wyznacza się liczbę chorych konieczną do udowodnienia różnicy w testach znamienności statystycznej. Zwraca jednak uwagę ważny fakt — porównywane są dwa ramiona przy założeniu, że jedynym elementem decydującym o istnieniu różnicy jest zastosowany lek lub metoda leczenia. Niestety, nikt nie jest w stanie udowodnić, że poza tym chorzy w obu ramionach są identyczni. Pozostaje wiara, że dzięki stratyfikacji i randomizacji faktycznie tak jest. W ten sposób jednak wartość naukowego dowodu matematycznego staje się zależna od tego, w co wierzymy. Każda nierzetelność w prowadzeniu badania, na każdym jego etapie, może podważyć wartość uzyskanego wyniku. A zatem nauka czy tylko wiara?

Aby uwiarygodnić wynik badania, stosuje się dalsze testy statystyczne. Jednym z nich jest tzw. iloraz ryzyka (HR, *hazard ratio*). Oblicza się go w ten sposób, że ryzyko zdarzenia w jednym ramieniu (liczba chorych ze zdarzeniem dzielona przez sumę chorych ze zdarzeniem i bez zdarzenia w danym ramieniu) dzieli się przez tak samo obliczone ryzyko w drugim ramieniu. Można to odnieść do ogólnej liczby chorych lub do poszczególnych kohort, utworzonych — przykładowo — ze względu na wiek, zaawansowanie choroby lub inne kryterium [3]. Na tej podstawie wnioskuje się, o ile nowy lek lub metoda redukuje ryzyko zdarzenia. Chodzi jednak nie o ryzyko zdarzenia u indywidualnego chorego, ale o ryzyko zdarzenia w grupie chorych z danego ramienia. Jest to więc to samo, co widać na wykresach, tylko inaczej pokazane. Zwróćmy uwagę, że kliniczna wartość współczynnika HR będzie zależała nie tylko od liczby zdarzeń, ale również od liczebności chorych w ramionach badania lub poszczególnych kohortach. Przykładowo, jeśli w ramieniu A są 3 zdarzenia na 10 chorych, a w ramieniu B 6 zdarzeń również na 10 chorych, to HR wyniesie  $0,3 : 0,6 = 0,5$ . Jeśli w innym badaniu w ramieniu A będą 3 zdarzenia na 100 chorych, a w ramieniu B 6 zdarzeń również na 100 chorych, to HR wyniesie  $0,03 : 0,06 = 0,5$ . Taki sam wynik uzyskamy przy odpowiednio 3 i 6 zdarzeniach na 1000, 10 000 lub więcej chorych w każdym z ramion. We wszystkich powyższych przykładach powiemy, że nastąpiła redukcja ryzyka zdarzenia o 50%. Ta sama wielkość tej redukcji będzie jednak miała zupełnie inne znaczenie kliniczne.

Kolejną analizą statystyczną jest tzw. drzewko (*forest plot*), które pokazuje wyniki w poszczególnych kohortach chorych (utworzonych według wieku, zaawansowania choroby i innych parametrów). Jeżeli w omawianej analizie stwierdza się różnice uzyskanego wyniku pomiędzy chorymi z różnych ramion, to najczęściej wnioskuje się, że wszyscy chorzy odnoszą korzyść z zastosowania nowego leku lub metody. Jest to oczywiście wniosek błędny. Wynik takiej analizy wskazuje jedynie na to, że chorzy którzy odnieśli korzyść z leczenia nowym lekiem lub metodą, należą do wszystkich lub prawie wszystkich kohort. W każdej kohorcie są zatem zarówno chorzy, którzy korzyść odnieśli, jak i ci, którzy takiej korzyści nie odnieśli. To niestety utrudnia wykrycie chorych, którzy faktycznie mogą odnieść korzyść z leczenia nowym lekiem lub metodą, przez co można wnioskować, że należy leczyć wszystkich, którzy spełniają kryteria, jakie przyjęto do kwalifikacji chorych do danego badania. Mają to udowodnić testy statystyczne, które jednak — bez względu na liczbę zastosowanych — nie mogą pokazać niczego więcej niż to, że w obu ramionach są zdarzenia i tylko w ramieniu badanym jest ich (najczęściej tylko trochę) mniej. Mnożenie testów statystycznych, które pokazują to samo w różny sposób, przypomina bardziej reklamę leku niż dowód naukowy.

### Czego brakuje w prezentacji wyniku?

W badaniu porównuje się częstość zdarzeń pomiędzy ramionami. Jednak podczas każdej oceny w każdym z ramion są chorzy, u których w danym czasie występuje lub nie jest odnotowane zdarzenie. Zauważmy, że przebieg krzywych wskazuje, że w każdym ramieniu znajdują się chorzy, u których stwierdza się oceniane zdarzenie już podczas pierwszych badań kontrolnych, i chorzy, u których zdarzenie nie występuje aż do zakończenia badania. Nie znamy decydujących o tym różnic pod względem charakterystyki chorych, bo przecież wszyscy spełniają te same kryteria kwalifikacji. Możemy przypuszczać, że wynika to z dopuszczenia do badania chorych cechujących się różnymi znanymi czynnikami ryzyka zdarzenia (stratyfikacja), ale hipoteza ta nie została poddana żadnej analizie. Można tylko przypuszczać, że zdarzenie będące punktem końcowym w pierwszej kolejności wystąpi u chorych, którzy już przy włączeniu do badania cechują się największym ryzykiem tego zdarzenia, np. z powodu zaawansowania choroby. To ta grupa znajdzie się w pierwszej połowie chorych z ocenianymi zdarzeniami, a więc zadecyduje o medianie. Czy tak jest w istocie — nie wiemy. Jeśli jednak byłoby to prawdą, to o wyniku każdego badania decydowałiby chorzy o największym ryzyku zdarzenia, a pozostali byłiby tylko rodzajem „dodatku” uzasadniającego zastosowanie nowego leku lub metody u wszystkich chorych spełniających kryteria włączenia, choć dla wielu

z nich może nie mieć znaczenia, w którym ramieniu się znajdują. Nie wiemy również, dlaczego przy ocenie dokonywanej w danym czasie zdarzenia ujawniają się w obu ramionach (choć w grupie badanej jest ich mniej). Jakie wspólne cechy mają ci chorzy z obu ramion? Jeśli mają jakąś wspólną cechę, to co powoduje różnicę w liczbie zdarzeń pomiędzy ramionami? Takie pytania można mnożyć, ale rzadko się zdarza, żeby szukano na nie odpowiedzi. Nie ulega jednak wątpliwości, że tylko w ten sposób można znaleźć chorych, którzy powinni być leczeni nowym lekiem lub nową metodą, bo tylko oni odniosą rzeczywistą korzyść z takiego leczenia.

## Podsumowanie

Wydaje się, że większą wagę przywiązuje się do ugruntowania powszechnego przekonania, że należy leczyć

wszystkich chorych spełniających kryteria kwalifikacji do badania, mimo że faktyczną korzyść odniesie niewielu z nich. Dla większości chorych uczestniczących w badaniu w obu ramionach czas do zdarzenia jest podobny.

W ten sposób badanie kliniczne staje się przede wszystkim metodą promowania leku lub metody, którą zastosowane analizy statystyczne mają uwiarygodnić.

## Piśmiennictwo

1. Kaplan EL, Meier P. Nonparametric estimation from incomplete observation. *Journal of the American Statistical Association*. 1958; 53(282): 457–481 (dostępny: <https://web.stanford.edu/~lutian/coursepdf/KMpaper.pdf>).
2. Green S, Benedetti J, Crowley J. *Clinical trials in oncology second edition*. Chapman and Hall/CRC. Boca Raton, London, New York, Washington DC. 2003: 30–37.
3. Green S, Benedetti J, Crowley J. *Clinical trials in oncology second edition*. Chapman and Hall/CRC. Boca Raton, London, New York, Washington DC. 2003: 37–39.