

Statystyka w praktyce hematologicznej

Statistics in hematological practice

Streszczenie

Wraz ze stale rosnącą liczbą badań w dziedzinie hematologii, znajomość metod statystycznych wykorzystywanych w analizie i interpretacji wyników stała się niezbędnym narzędziem pracy klinicystów. W artykule omówiono najczęściej stosowane testy statystyczne oraz zdefiniowano punkty końcowe stosowane podczas raportowania rezultatów badań klinicznych. Testy statystyczne funkcjonują na zasadzie testowania hipotez. Odrzucenie lub nieodrzućenie danej hipotezy zależy od wybranego poziomu istotności oraz wyliczonej wartości p . Z kolei otrzymany przedział ufności wskazuje na wielkość efektu i precyzję oszacowania. W hematologii głównie raportowanymi punktami końcowymi są: całkowite przeżycie (ang. *overall survival* – OS), przeżycie wolne od progresji (ang. *progression-free survival* – PFS), odpowiedź na leczenie (ang. *overall response rate* – ORR) oraz coraz częściej spotykana, ocena terapii raportowana przez pacjenta (*patient reported outcome* – PRO). Biorąc pod uwagę kierunek rozwoju medycyny, znajomość biostatystyki jest niezbędna w podejmowaniu decyzji terapeutycznych, a także ocenianiu, interpretowaniu i raportowaniu wyników przeprowadzonych badań.

Article history:

Received: 02.06.2018

Accepted: 11.10.2018

Iga Andrasiak¹,
Tomasz Wróbel²¹ Dział Badawczy WroMedica, Wrocław,
Polska² Katedra i Klinika Hematologii,
Nowotworów Krwi i Transplantacji Szpiku,
Uniwersytet Medyczny im. Piastów Śląskich
we Wrocławiu, Polska

Abstract

With the ever increasing number of articles in the field of hematology, the knowledge of applied statistical methods for analysis and interpretation of results, became a necessary skill for clinicians. Article describes the most common statistical tests and defines endpoints applied to clinical trial reports. Statistical analysis is based on hypothesis testing. Reject or fail to reject hypothesis depends on a set significance level and the calculated p -value. Furthermore, the obtained confidence interval shows the size effect and the precision of evaluation. The prime reported endpoints in hematology are: overall survival (OS), progression-free survival (PFS), overall response rate (ORR) and increasingly used, patient reported outcome (PRO). Given the direction of medicine development, knowledge of biostatistics is necessary in order to make therapeutic decisions, as well as to evaluate, interpret and report the study results.

© 2018 Polish Society of Hematology and Transfusion Medicine, Insitute of Hematology and Transfusion Medicine. All rights reserved.

Słowa kluczowe:

przedział ufności, p -wartość, punkty końcowe, analiza przeżycia, całkowite przeżycie

Keywords:

confidence interval, p -value, endpoints, survival analysis, overall survival

Wstęp

Elementy statystyki są nieodłączną częścią współczesnej medycyny. Klinicyści na co dzień spotykają się z wieloma pojęciami statystycznymi, tj. średnia, mediana, istotność statystyczna czy przedział ufności. Medycyna oparta na faktach (ang. *evidence-based medicine* – EBM) optymalizuje proces decyzyjny poprzez użycie wiarygodnych danych naukowych w codziennej praktyce klinicznej. Znajomość biostatystyki jest niezbędna, aby zapewnić pacjentom właściwą opiekę, podejmować optymalne decyzje terapeutyczne, poprawnie oceniać, interpretować i raportować wyniki przeprowadzonych badań. Celem artykułu jest omówienie pojęć z zakresu statystyki, spotykanych w artykułach dotyczących schorzeń układu kriotwórczego. Usystematyzowanie wiedzy z zakresu biostatystyki pomoże w interpretacji i krytycznej ocenie otrzymanych wyników, natomiast jednorodność stosowanych metod pozwoli na dalsze porównania powstających artykułów, m.in. z użyciem metaanaliz.

Opis zmiennych

W publikacjach hematologicznych dominuje opis danych poprzez wartości liczbowe. Zmienne jakościowe, niemierzalne, czyli dane kategoryjne, np. płeć, typ nowotworu, są opisywane jako ilość

danych obserwacji w próbie oraz ich częstość (stosunek liczby obserwacji do liczebności próby). Zmienne ilościowe, przyjmujące wartości numeryczne, opisywane są za pomocą miary centralnego rozmieszczenia wartości zmiennej, wyrażonej jako:

- średnia arytmetyczna próby, czyli iloraz sumy obserwacji przez całkowitą ilość obserwacji,
- mediana, czyli wartość środkowa uszeregowanego zbioru danych.

Natomiast miarami zmienności szeregu obserwacji są:

- odchylenie standardowe (SD), definiowane jako pierwiastek kwadratowy ilorazu sumy kwadratów odchyleń wartości zmiennej (X) od średniej arytmetycznej zbioru (\bar{X}) przez ilość osób w próbie (N) pomniejszonej o 1:

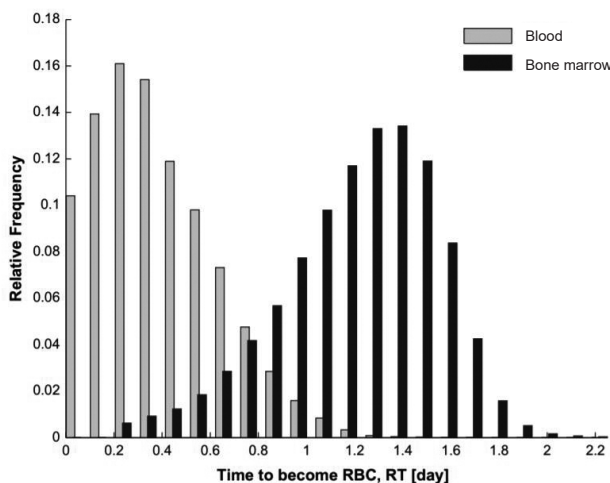
$$SD = \sqrt{\frac{\sum_{i=1}^n (X - \bar{X})^2}{N - 1}}$$

- rozstęp, czyli różnica między obserwacją maksymalną a minimalną; w praktyce podawany jest zakres wartości od najmniejszej do największej.

Im zbiór danych będzie bardziej zróżnicowany, tym wartość odchylenia standardowego będzie większa. Średnia podawana jest

* Adres do korespondencji: Iga Andrasiak, Dział Badawczy WroMedica, ul. A. Mickiewicza 91, 51-685 Wrocław, Polska; e-mail: igaandrasiak@gmail.com

wraz z wartością SD, natomiast przy medianie wskazywany jest zakres wartości zmiennej. Trzeba mieć na uwadze, że jakkolwiek skośność układu będzie miała znaczący wpływ na wartość średnią. Inaczej będzie z medianą, która jest odporna na niesymetryczność danych oraz na wartości odstające. Dlatego warto najpierw przyrzeć się rozkładowi zmiennej, a następnie wybrać miarę, która będzie najlepiej reprezentować daną próbę. Przydatnym, lecz stosunkowo rzadko spotykanym narzędziem w artykułach medycznych jest współczynnik skośności, będący miarą asymetrii rozkładu. Wskazuje on, czy większość obserwacji znajduje się w pobliżu wartości średniej – wtedy mówimy o rozkładzie symetrycznym (współczynnik skośności wynosi 0), czy też powyżej lub poniżej wartości średniej, dając współczynnik skośności odpowiednio ujemny lub dodatni. Rozkłady skośne zaobserwowali Wiczling i Krzyżanski [1] podczas analizy względnego czasu przebywania retikulocytów we krwi oraz w szpiku kostnym u szczurów za pomocą cytometrii przepływowej (Ryc. 1). Współczynnik skośności dla rozkładu zmiennej we krwi wynosił 0,15, zatem zmienna była prawostronnie asymetryczna, oraz -0,16 w szpiku kostnym, wskazując na asymetrię lewostronną. W asymetrii prawostronnej prawe ramię rozkładu jest wydłużone, a większość obserwacji znajduje się poniżej średniej. Odwrotnie jest w asymetrii lewostronnej, gdy wydłuża się lewe ramię rozkładu i dominują obserwacje powyżej wartości średniej. W przytoczonym artykule miarą centralnego rozmieszczenia zmiennej we krwi, jak i w szpiku kostnym jest mediana wynosząca odpowiednio 0,39 i 1,30 dnia, która ze względu na skośność rozkładu różni się od wartości średniej wynoszącej odpowiednio 0,43 i 1,36 dnia.



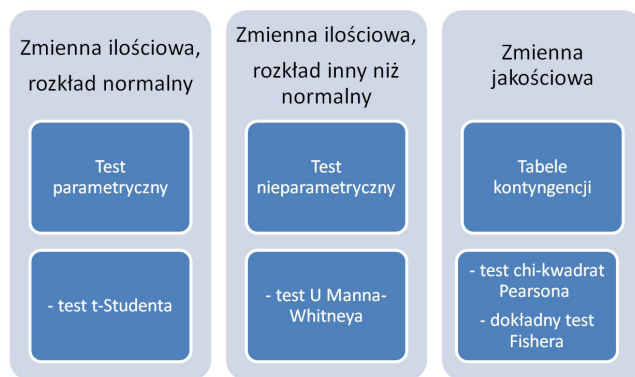
Ryc. 1. Histogramy rozkładu czasu obecności retikulocytów we krwi i w szpiku kostnym zaraportowane przez Wiczling i Krzyżanski [1]

Fig. 1. Histograms of the age distribution for homeostatic blood and bone marrow reticulocytes reported by Wiczling and Krzyżanski [1]

Rozkład normalny i podstawowe testy statystyczne

Powszechnie uważa się, że rozkład normalny, zwany rozkładem Gaussa, jest najczęściej spotykanym rozkładem w publikacjach

naukowych. Wykres funkcji wielu zmiennych losowych przyjmuje charakterystyczny kształt dzwonu. W artykułach z zakresu hematologii dominuje weryfikacja normalności rozkładu z użyciem testu Shapiro-Wilka oraz testu Kołmogorowa-Smirnowa. Przy dalszej analizie danych rozkład Gaussa pozwoli autorom skorzystać z wielu popularnych i wygodnych testów statystycznych – testów parametrycznych, do których należą test t-Studenta czy analiza wariancji (ANOVA). W publikacji, w sekcji *Metody*, powinna znaleźć się informacja o sposobie weryfikacji rozkładu zmiennej. Jej pominięcie sugeruje czytelnikom, że normalność rozkładu nie została sprawdzona, a zatem zarówno zasadność zastosowania danego testu statystycznego, jak i otrzymane wyniki mogą być niepoprawne. Czy niespełnienie założenia o rozkładzie normalnym wyklucza dalszą analizę danych? Oczywiście odpowiedź brzmi: nie. W przypadku braku rozkładu normalnego skutecznie można skorzystać z testów nieparametrycznych, do których zalicza się test U Manna-Whitneya czy też test Wilcozona dla par obserwacji. W klasycznej sytuacji, w której zestawiono 2 grupy badanych (interwencja vs kontrola albo interwencja A vs interwencja B), porównania dokonuje się za pomocą testu t-Studenta (gdy spełnione są założenia testu parametrycznego), testu U Manna-Whitneya (test nieparametryczny) lub testów chi-kwadrat Pearsona albo dokładnego testu Fishera, gdy rozpatrywane są zmienne kategoryjne (Ryc. 2).



Ryc. 2. Metody statystyczne stosowane przy testowaniu hipotez dla 2 grup pacjentów

Fig. 2. Statistical methods used to hypothesis testing of 2 groups of patients

Zmienne kategoryjne graficznie przedstawia się w tabeli kontyngencji, zwanej również tabelą krzyżową lub tabelą wielodzzielczą. Najprostszą tabelą kontyngencji jest tabela dwudzzielcza 2 x 2, w której obie zmienne są dychotomiczne, czyli przyjmują dwie wartości. Każdy pacjent przypisywany jest tylko do jednej, konkretnej komórki, a suma pacjentów w komórkach jest zliczana. Test statystyczny polega na odrzuceniu hipotezy zerowej (H_0), która mówi, że nie ma istotnej różnicy między liczebnościami obserwowanymi (O) w komórkach a liczebnościami oczekiwanymi (E). Dodatkowo, w przypadku tabeli dwudzzielczej test chi-kwadrat wymaga, aby ilość obserwacji obserwowanych wynosiła przynajmniej 5 w każdej komórce. Natomiast w przypadku mniejszych prób zaleca się stosowanie dokładnego

testu Fishera. Oba testy często wykorzystywane są przy porównywaniu danych demograficznych, ilości odpowiedzi na leczenie lub też częstości występowania działań niepożądanych pomiędzy grupami pacjentów. Przykładem zastosowania testu chi-kwadrat jest analiza częstości wystąpienia neutropenii według schematu terapii R-CHOP (rytuksymab, cyklofosfamid, doksorubicyna, winkrystyna, prednizolon) vs R-HDS (rytuksymab z sekwencjami chemioterapii w wysokich dawkach) [2], przedstawiona na rycinie 3. Do obliczenia liczebności oczekiwanych zastosowano formułę:

$$E = \frac{(\text{suma rzędu}) \cdot (\text{suma kolumny})}{(\text{suma całkowita})}$$

Wzór testu chi-kwadrat ma postać:

$$\chi^2 = \sum_{i=1}^{w \times k} \frac{(O_i - E_i)^2}{E_i}$$

gdzie w (wiersze) i k (kolumny) to wymiary tabeli.

P-wartość a przedział ufności

Każde badanie naukowe ma na celu sformułowanie, a następnie odrzucenie bądź przyjęcie konkretnej hipotezy. Weryfikacja hipotez jest domeną statystyki, a odbywa się za pomocą odpowiednio dobranych testów statystycznych. Każdy test rozpoczyna się od ustalenia hipotezy zerowej H_0 oraz hipotezy alternatywnej H_1 . Testy statystyczne z reguły zakładają w H_0 brak różnic (np. pomiędzy skutecznością leków), podczas gdy klinicytom zależy właśnie na wykazaniu tych różnic, a zatem odrzuceniu H_0 i przyjęciu H_1 . O odrzuceniu H_0 decyduje obliczona wartość p. Jeśli wartość p jest niższa od przyjętego poziomu istotności (zwykle wartość graniczna wynosi 0,05), to odrzuca się H_0 i przyjmuje H_1 . Natomiast gdy wartość p przewyższa krytyczny poziom istotności, nie ma podstaw

Czym jest przyjęty poziom istotności?

Jest to błąd pierwszego rodzaju, który jesteśmy skłonni zaakceptować. Polega on na odrzuceniu H_0 , która w rzeczywistości jest prawdziwa. Wartości graniczna $p=0,05$ wskazuje, że liczymy się z 5 pomyłkami na 100 przeprowadzonych prób.

do odrzucenia H_0 . Druga sytuacja niekoniecznie oznacza, że skuteczność terapii jest identyczna. To konkretna próba nie wykazuje istotnej statystycznie różnicy, co może być spowodowane zbyt niską mocą statystyczną.

Poziom istotności przyjmowany jest w zależności od natury badania i decyzji zespołu badawczego. W badaniach hematologicznych najczęściej spotyka się wartość odcięcia na poziomie 0,05; zdecydowanie rzadziej 0,01. Autorzy powinni podawać przyjęty poziom istotności w sekcji *Metody*. Należy w niej także zawrzeć przyjęty poziom pogranicza istotności (*ang. borderline*), jeśli z takiego terminu korzysta się w dalszej części artykułu. W przeciwnym wypadku czytelnik nie ma przedstawionych jasnych wytycznych klasyfikacji i nie wie, czy dany wynik jest dla autorów istotny, na pograniczu istotności, czy może już nieistotny statystycznie.

Wyliczona p-wartość pozwala na przyjęcie lub odrzucenie danej hipotezy. Natomiast przedział ufności (CI) ukazuje wielkość efektu i precyzję oszacowania. Dlatego też w ostatnich czasach kładzie się duży nacisk na raportowanie obu tych wyników [3]. Koncepcję przedziału ufności opracował polski matematyk i statystyk, Jerzy Sława-Neyman [4]. Bazując na wyniku próby, przedział ufności pozwala autorom estymować wielkość efektu na całą populację. Dodatkowo dolna i górna granica przedziału znacząco ułatwiają

		Neutropenia stopnia III lub IV podczas terapii		
		Tak	Nie	Suma
Terapia	R-CHOP	42 [71,12]	80 [50,88]	122
	R-HDS	95 [65,88]	18 [47,12]	113
Suma		137	98	235

Ryc. 3. Analiza częstości wystąpienia neutropenii w terapii wg schematu R-CHOP vs R-HDS. W komórkach podano liczebność obserwacji O wraz z liczebnością oczekiwaną [E]. Wartość statystyki wynosi $\chi^2 = 59,47$, $p < 0,001$. Dane pochodzą z publikacji Cortelazzo i wsp. [2]. R-CHOP – rytuksymab, cyklofosfamid, doksorubicyna, winkrystyna, prednizolon; R-HDS – rytuksymab z sekwencjami chemioterapii w wysokich dawkach

Fig. 3. Frequency analysis of neutropenia occurrence in the R-CHOP arm vs R-HDS arm. Observed value O and expected value [E] are provided in each cell. Value of $\chi^2 = 59,47$, $p < 0,001$. Data reported by Cortelazzo et al. [2]. R-CHOP – rituximab, cyclophosphamide, doxorubicin, vincristine, prednisone; R-HDS – rituximab plus high-dose sequential chemotherapy

badaczom interpretację otrzymanych wyników. Można się spodziewać, że badania przeprowadzone na dużej próbie pacjentów będą miały węższy przedział ufności, czyli większą precyzję oszacowania niż małe próby. Szeroki przedział ufności oznacza duże odchylenie wartości cechy oszacowanej w badaniu od faktycznej wartości cechy w populacji, a zatem małą wiarygodność wyniku. Wartość CI podaje się zawsze wraz z wybranym poziomem ufności (1- α), który zazwyczaj wynosi 95%. Oznacza on, że na 100 losowych prób z populacji, 95 przedziałów będzie zawierało estymowany parametr.

Punkty końcowe

Punkt końcowy definiowany jest jako zdarzenie lub rezultat, którego częstość pozwala wnioskować o korzyści stosowania terapii przez pacjenta. Porównanie odsetka pacjentów odpowiadających na leczenie (ORR) wiąże się z opisywanym uprzednio testem chi-kwadrat. Punkty końcowe będące zmienną ilościową, na przykład kwestionariusze jakości życia, mogą być analizowane testem t-Studenta lub testem U Manna-Whitneya. Co ciekawe, niektóre dziedziny medycyny precyzyjnie identyfikują punkty końcowe, które są konsekwentnie powtarzane w kolejnych artykułach. W hematologii decyzja o wyborze punktów końcowych jest bardziej skomplikowana. Należy bowiem uwzględnić, że nawroty choroby są ściśle związane ze skutecznością leczenia, a także z wybranym czasem obserwacji. Naprzeciw oczekiwaniom klinicystów wychodzi biostatystyka z analizą przeżycia. Analiza przeżycia pozwala porównać czas, po którym nastąpi oczekiwane zdarzenie, to jest punkt końcowy badania. Zatem badany jest rozkład zdarzeń w czasie, a nie sam fakt ich wystąpienia. Ponieważ pacjenci są włączani do badania w różnym czasie, moment wystąpienia punktu końcowego wyznaczany jest indywidualnie, co jest dużą zaletą tej metody. Istotnym pojęciem w analizie przeżycia jest hazard względny (ang. *hazard ratio* – HR), powszechnie podawany z 95% CI. HR to stosunek ryzyka wystąpienia punktu końcowego w dowolnym punkcie czasowym. Innymi słowy, HR to chwilowy potencjał pojawienia się konkretnego zdarzenia w czasie, łatwy do zobrazowania jako analogia chwilowych odczytów prędkości na prędkościomierzu w samochodzie.

Najstarszą metodą opisu funkcji przeżycia jest estymacja poprzez tablice przeżycia. Obecnie niepraktykowana, ponieważ wymaga arbitralnego grupowania czasu obserwacji. Z reguły można spotkać się z analizą przeżycia metodą Kaplana-Meiera. Bazuje ona na logicznym założeniu, że aby przeżyć 5 miesięcy, trzeba najpierw przeżyć 4 wcześniejsze. Wyznaczany estymator jest zatem iloczynem kolejnych prawdopodobieństw warunkowych. Celem badań medycznych jest porównanie uzyskanych krzywych przeżycia między grupami pacjentów, co można zrobić za pomocą, m.in.: testu logrank, testu Coxa-Mantela, testu F Coxa czy testu Wilcoxa według Gehana. W wymienionych analizach testowana jest hipoteza zerowa o braku różnic funkcji przeżycia pomiędzy grupami. Najczęściej stosowanym testem analizy przeżycia pozostaje test logrank [5] – test nieparametryczny i odporny na obserwacje odstające.

Badacze często zakładają, że istnieją dodatkowe zmienne, które wpływają na czas przeżycia. Przykładem może być raport Decaux i wsp. [6], w którym wykazano, że na czas przeżycia po zastosowaniu talidomidu w zaawansowanym szpiczaku mnogim wpływają:

oporność na wcześniejszą terapię, stan sprawności pacjenta, poziom β -mikroglobuliny oraz płytek, a także brak odpowiedzi na leczenie w dniu 60. Mówimy wtedy o wielowymiarowej analizie przeżycia. Istnieje wiele dostępnych modeli do wykorzystania, w tym model regresji liniowej, model regresji log-normalnej czy model wykładniczy. W praktyce onkologicznej prawie zawsze spotyka się nieparametryczny model proporcjonalnego hazardu Coxa, definiowanego wzorem:

$$h(t; x_1, x_2, \dots, x_n) = h_{0(t)} \times \exp\{b_1x_1 + b_2x_2 + \dots + b_nx_n\}$$

gdzie: h to wyraz wolny, zwany hazardem podstawowym, natomiast $h(t; x_1, x_2, \dots, x_n)$ to hazard zmiennych towarzyszących x_1, x_2, \dots, x_n w czasie t . W publikacjach zwykle podawany jest wynik analizy przeżycia obejmujący okres 1 roku oraz 2 lat.

Nazwa „analiza przeżycia” może być mylnie interpretowana jako metoda oceny tylko i wyłącznie całkowitego przeżycia (ang. *overall survival* – OS). W rzeczywistości służy ona badaczom do oceny wszelkich punktów końcowych związanych z czasem, na przykład czasu przeżycia wolnego od progresji (ang. *progression-free survival* – PFS). W takich sytuacjach przedmiotem analizy nie jest zgon, a dany punkt końcowy, tutaj: progresja choroby lub zgon. Zestawienie najczęściej raportowanych punktów końcowych w onkologii, a w szczególności w hematologii przedstawiono w tabeli I.

Całkowite przeżycie to powszechnie akceptowany oraz najczęściej oznaczany punkt końcowy w całej onkologii. Jest to przedział czasu od momentu randomizacji lub włączenia pacjenta do badania aż do wystąpienia zgonu z jakiegokolwiek przyczyny. Wyznaczanie OS ma sporo zalet – jest łatwe i precyzyjne w pomiarze, a dzięki swojej popularności badacze mogą bez problemu porównywać OS pomiędzy konkretnymi terapiami. Niestety, na wartość OS wpływają również inne choroby i terapie, nie tylko te związane z chorobą nowotworową. Ponadto, aby wykazać istotną statystycznie różnicę, potrzeba dużej populacji pacjentów oraz stosunkowo długiego czasu obserwacji.

OS należy do tzw. twardych punktów końcowych. Obecnie autorzy mają możliwość skorzystania z surogatów, czyli wygodniejszych do weryfikacji, zastępczych punktów końcowych. Te zamienniki stosuje się, aby ograniczyć koszty związane z badaniem. Pozwalają przewidzieć, co będzie się działo z pacjentem i oszacować skuteczność leczenia, wykorzystując mniejszą wielkość próby oraz krótszy czas obserwacji. W randomizowanych badaniach klinicznych coraz częściej wykorzystuje się surogaty jako pierwszorzędowe punkty końcowe [8]. Jako przykład może posłużyć PFS, który stosuje się jako surogat OS. PFS definiowany jako czas od momentu włączenia do badania (np. randomizacji, pierwszego podania leku) aż do wystąpienia progresji choroby lub zgonu. Jako surogat OS, PFS będzie mógł wykazać istotność statystyczną w mniejszej grupie pacjentów w stosunkowo krótszym czasie. Należy pamiętać, że aby PFS dał wiarygodne rezultaty, konieczna jest obiektywna ocena progresji choroby. Ponadto nie zawsze opóźnienie progresji wiąże się z wydłużeniem czasu przeżycia. Przy ocenie korzyści terapii warto rozważyć nie tylko PFS, lecz także toksyczność terapii oraz raportowaną przez pacjenta jakość życia. Obecnie PFS to powszechnie wykorzystywany punkt końcowy występujący w artykułach hematologicznych, akceptowany przez agencje regulatorowe: FDA (*Food and Drug Administration*), EMA (*European*

Tabela I. Punkty końcowe stosowane w onkologii [7]

Table I. Endpoints used in oncology [7]

Punkt końcowy	Charakterystyka
Całkowite przeżycie (ang. <i>overall survival, OS</i>)	Czas pomiędzy punktem początkowym (np. randomizacją) a zgonem.
Czas przeżycia wolny od progresji (ang. <i>progression-free survival, PFS</i>)	Czas pomiędzy punktem początkowym (np. randomizacją) a progresją choroby lub zgonem.
Czas do progresji (ang. <i>time to progression, TTP</i>)	Czas pomiędzy punktem początkowym (np. randomizacją) a progresją choroby.
Czas przeżycia wolny od choroby (ang. <i>disease-free survival, DFS</i>)	Czas pomiędzy punktem początkowym obserwacji a nawrotem choroby lub zgonem.
Czas do niepowodzenia leczenia (ang. <i>time to treatment failure, TTF</i>)	Czas pomiędzy punktem początkowym (np. randomizacją) a dyskontynuacją leczenia.
Odpowiedź na leczenie (ang. <i>overall response rate, ORR</i>)	Odsetek odpowiedzi na leczenie.

Medicines Agency) oraz PDMA (*Pharmaceuticals and Medical Devices Agency*) [9].

Pokrewnym punktem końcowym do PFS jest czas do progresji (ang. *time to progression – TTP*). W przypadku TTP wyznaczany jest czas od włączenia do badania do progresji choroby. W przeciwieństwie do PFS, TTP nie uwzględnia zgonów pacjenta, z wyjątkiem śmierci w następstwie progresji choroby nowotworowej. W związku z tym uznaje się, że TTP jest gorszym punktem zastępczym OS niż PFS. Jednak w sytuacjach, gdy badacze mogą założyć, że większość zgonów jest niezwiązanych z chorobą podstawową pacjenta, TTP może lepiej odzwierciedlać faktyczną skuteczność terapii.

Czas wolny od choroby (ang. *disease-free survival – DFS*) często jest utożsamiany z PFS. I faktycznie, w konkretnych przypadkach ich wartości są identyczne, jednak nie zawsze. DFS jest bowiem definiowany jako czas od włączenia do badania do momentu nawrotu choroby lub śmierci z jakiegokolwiek przyczyny. Jednak o nawrocie choroby można mówić tylko i jedynie w sytuacji, gdy wszyscy lub duży odsetek pacjentów uzyskał odpowiedź na leczenie. Dlatego w artykułach onkologicznych wykorzystuje się DFS jako surogat OS w wybranych wskazaniach, na przykład w badaniach raka piersi [10] czy raka okrężnicy [11]. W hematologii jest stosowany raczej przy analizie skuteczności leczenia podtrzymującego [12] lub weryfikacji nawrotów choroby [13].

Czas do niepowodzenia leczenia (ang. *time to treatment failure – TTF*) jest to złożony punkt końcowy. Uwzględnia czas od momentu włączenia do badania (np. podania leku, randomizacji) do dyskontynuacji leczenia z jakiegokolwiek powodu (włączając: progresję choroby, toksyczność leku, śmierć, wycofanie z badania).

TTF obrazuje rzeczywisty czas trwania leczenia. Nie jest on jednak rekomendowanym pierwszorzędowym punktem końcowym, ponieważ badacze nie są w stanie wyodrębnić skuteczności leczenia od innych zmiennych, na przykład toksyczności. TTF jest w artykułach podawany informacyjnie lub jako drugorzędowy punkt końcowy. Odpowiedź na leczenie (ang. *overall response rate – ORR*) definiowana jest jako suma odsetka całkowitych odpowiedzi (CR) oraz częściowych odpowiedzi (PR) na leczenie. W skład ORR badacze nie powinni włączać odsetka pacjentów ze stabilną chorobą, gdyż pożądanym efektem terapeutycznym w onkologii i hematologii jest zmniejszenie guza, a nie utrzymanie jego rozmiaru. Preferowane jest, aby dla wyznaczania CR, PR, odsetka pacjentów ze stabilną chorobą i progresją, przyjmowane były standaryzowane kryteria dla danej jednostki chorobowej, np. w przypadku guzów litych bazujące na wytycznych RECIST [14]. Należy też pamiętać, że obserwowana ocena będzie ściśle zależeć od punktu czasowego pomiaru. Innymi punktami końcowymi stosowanymi w artykułach z zakresu hematologii są:

- Ocena terapii raportowana przez pacjenta (ang. *patient reported outcome – PRO*); twardy punkt końcowy, definiowany jako określenie stanu zdrowia bezpośrednio przez pacjenta, bez ingerencji lub interpretacji lekarza albo osób trzecich. W ostatnich czasach PRO jest coraz popularniejszy i bardziej ceniony przez naukowców [15].
- Czas wolny od zdarzeń (ang. *event-free survival – EFS*) to czas od włączenia do badania (pierwszego podania leku, randomizacji) do zaistnienia istotnego zdarzenia, włączając: progresję choroby, śmierć, dyskontynuację leczenia. Zakończenie terapii

może być związane z jej toksycznością, rezygnacją pacjenta z udziału w badaniu lub decyzją o zmianie terapii mimo braku progresji. Punkt końcowy nie jest preferowany ze względu na nakładanie się informacji o skuteczności, toksyczności i wycofaniu z leczenia.

Opublikowane rezultaty z dokładnie opisaną metodologią oraz spełniające konkretne kryteria można powtórnie wykorzystać, na przykład w porównaniu pośrednim lub metaanalizie. Porównanie pośrednie pozwala zestawić skuteczność wcześniej nieporównywanych związków, które były analizowane w niezależnych randomizowanych badaniach z tym samym komparatorem. Na przykład wykorzystując badania: ibrutynib vs chlorambucyl oraz bendamustyna vs chlorambucyl, porównano pośrednio ibrutynib z bendamustyną, wykazując, że ibrutynib istotnie zwiększa OS u pacjentów z przewlekłą białaczką limfocytową [16]. Z kolei metaanaliza jest syntezą wyników uzyskanych w niezależnych badaniach, której wyniki dają szerszy pogląd na badane zagadnienie niż analizowanie pojedynczych raportów. Obrazuje to artykuł, w którym udowodniono, że w chłoniaku indolentnym schemat podawania bortezomibu co 2 tygodnie jest skuteczniejszy niż co tydzień, przy podobnej toksyczności [17].

Wnioski

Wiedza z zakresu statystyki jest dla hematologów kluczowa przy projektowaniu, interpretacji oraz opracowywaniu wyników badań. Onkologia jest dynamicznie rozwijającą się dziedziną medycyny, ze stale rosnącą ilością artykułów i projektów, zatem znajomość zwyczajowo stosowanych metod statystycznych stała się niezbędnym

narzędziem pracy. Przy analizie wyników badań należy stosować testy statystyczne odpowiednio do badanych zmiennych, jak i efektu, który chce się wykazać. Powinno się zawsze weryfikować założenia oraz pamiętać, że zastosowanie jednolitej metodologii w artykułach z danej dziedziny ułatwi późniejsze porównanie wyników. Sekcja *Metody* nie powinna być najkrótszym fragmentem publikacji. Czytelnik musi być poinformowany, w jaki sposób otrzymano wyniki. Ponadto precyzyjny opis zastosowanych metod statystycznych, sposobu definiowania punktów końcowych czy też progresji choroby potwierdza wiarygodność otrzymanych rezultatów i wniosków. Pozwala też w przyszłości przeprowadzić analogiczne badanie albo powtórnie wykorzystać otrzymane wyniki, np. w metaanalizie.

Wkład autorów/Author's contributions

Według kolejności.

Konflikt interesu/Conflict of interest

Nie występuje.

Finansowanie/Financial support

Nie występuje.

Etyka/Ethics

Treści przedstawione w artykule są zgodne z zasadami Deklaracji Helsińskiej, dyrektywami EU oraz ujednoliconymi wymaganiami dla czasopism biomedycznych.

Piśmiennictwo/References

- [1] Wiczling P, Krzyżanski W. Flow cytometric assessment of homeostatic aging of reticulocytes in rats. *Exp Hematol* 2008;36(2):119–27. doi: 10.1016/j.exphem.2007.09.002.
- [2] Cortelazzo BS, Tarella C, Gianni AM, et al. Randomized trial comparing R-CHOP versus high-dose sequential chemotherapy in high-risk patients with diffuse large B-cell lymphomas. *J Clin Oncol* 2016; 34(33):4015–4022.
- [3] du Prel JB, Hommel G, Röhrig B, Blettner M. Confidence interval or P value?: Part 4 of a series on evaluation of scientific publications. *Dtsch Arztebl Int* 2009;106(19):335–9.
- [4] Kendall DG, Bartlett MS, Page TL. Jerzy Neyman. 16 April 1894–5 August 1981. *Biogr Mem Fell R Soc* 1982;28:379–412.
- [5] Fendler W, Chałubińska J, Młynarski W. Techniki analizy przeżycia stosowane w onkologii – założenia, metodyka i typowe problemy interpretacyjne. *Onkologia w Praktyce Klinicznej* 2011;7(2):89–101.
- [6] Decaux O, Renault A, Sébille V, et al. Predictive factors of survival after thalidomide therapy in advanced multiple myeloma: long-term follow-up of a prospective multicenter nonrandomized phase II study in 120 patients. *Clin Lymphoma Myeloma Leuk* 2012;12(6):418–22.
- [7] U.S. Department of Health and Human Services Food and Drug Administration Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER). Guidance for industry: clinical trial endpoints for the approval of cancer drugs and biologics. 2007; <https://www.fda.gov/downloads/Drugs/.../Guidances/ucm071590.pdf>.
- [8] Kemp R, Prasad V. Surrogate endpoints in oncology: when are they acceptable for regulatory and clinical decisions, and are they currently overused? *BMC Med* 2017;15(1):134.
- [9] Schmitter S, Brock E, Holzerny P, Günzel CA, Ruckdäschel S. Regulatory Agencies' Perspective on "Progression-Free Survival" (PFS), Poster-No. PCN289, ISPOR 19th Annual European Congress, 02/11/2016, Vienna.
- [10] Ishikawa T, Akazawa K, Hasegawa Y, et al. Survival outcomes of neoadjuvant chemotherapy with zoledronic acid for HER2-negative breast cancer. *J Surg Res* 2017;220:46–51.
- [11] Minami K, Morita M, Emi Y, et al. Final report of KSCC0803: feasibility study of capecitabine as adjuvant chemotherapy for stage III colon cancer in Japan. *Int J Clin Oncol* 2017;22(3):505–510.
- [12] Blum W, Sanford BL, Klisovic R, et al. Maintenance therapy with decitabine in younger adults with acute myeloid leukemia in first remission: a phase 2 Cancer and Leukemia Group B Study (CALGB 10503). *Leukemia* 2017; 31(1):34–39.
- [13] Desjonquères A, Chevallier P, Thomas X, et al. Acute lymphoblastic leukemia relapsing after first-line pediatric-inspired therapy: a retrospective GRAALL study. *Blood Cancer J* 2016;6(12):e504.

- [14] Eisenhauer EA, Therasse P, Bogaerts J. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer* 2009;45(2):228–47.
- [15] LeBlanc TW, Abernethy AP. Patient-reported outcomes in cancer care—hearing the patient voice at greater volume. *Nat Rev Clin Oncol* 2017;14(12):763–772.
- [16] Andrasiak I, Rybka J, Knopinska-Posluszny W, Wrobel T. Efficacy and Safety of Bendamustine and Ibrutinib in Previously Untreated Patients With Chronic Lymphocytic Leukemia: Indirect Comparison. *Clin Lymphoma Myeloma Leuk* 2017;17(5):268–273.
- [17] Yuan T, Zhang F, Yao QM, Liu YX, Zhu XJ, Wang X. Weekly versus biweekly bortezomib given in patients with indolent non-Hodgkin lymphoma: A meta-analysis. *PLoS One* 2017;12(5):e0177950.