

Wojciech Fendler<sup>1</sup>, Justyna Chałubińska<sup>2</sup>, Wojciech Młynarski<sup>1</sup>

<sup>1</sup>Klinika Pediatrii, Onkologii, Hematologii i Diabetologii Uniwersytetu Medycznego w Łodzi

<sup>2</sup>Zakład Teleradioterapii, Regionalny Ośrodek Onkologiczny w Łodzi

# Techniki analizy przeżycia stosowane w onkologii — założenia, metodyka i typowe problemy interpretacyjne

Methods of survival analysis applied in oncology  
— assumptions, methods and common pitfalls

## Adres do korespondencji:

Dr med. Wojciech Fendler  
Klinika Pediatrii, Onkologii, Hematologii  
i Diabetologii  
ul. Sporna 36/50, 91-738 Łódź  
Tel: +48 (42) 617 77 50  
Faks: +48 (42) 617 77 98  
email: Fendler.wojciech@gmail.com

## STRESZCZENIE

Analiza przeżycia stanowi fundament analityczny badań nad ryzykiem zgonu z powodu choroby nowotworowej lub jej progresji. Techniki analityczne stosowane w tym celu uwzględniają niepewność co do wystąpienia punktu końcowego w czasie. Przerwanie obserwacji pacjenta może następować z kilku powodów — zakończenia badania, niechęci do dalszego uczestnictwa w nim, konieczności zmiany terapii ze względu na czynniki kliniczne niezwiązane z punktem końcowym itp. Główną zaletą tych metod analizy jest możliwość szacowania indywidualnego ryzyka (hazardu) wystąpienia punktu końcowego w czasie, a w konsekwencji przewidywania oczekiwanego czasu przeżycia w zależności od różnorodnych czynników związanych z procesem chorobowym lub stosowaną terapią. Ze względu na olbrzymią i stale rosnącą liczbę artykułów naukowych z zakresu onkologii wykorzystujących analizy przeżycia do wykazania efektu biologicznego terapii, leku lub innej interwencji znajomość założeń i interpretacji powszechnie stosowanych metod powoli staje się nieodzowna. Wśród metod analitycznych używanych w tym celu są nieparametryczne analizy jednoczynnikowe (test log-rank i inne), modele wieloczynnikowe (modele proporcjonalnego i addytywnego ryzyka), sztuczne sieci neuronowe i drzewa klasyfikacyjno-regresyjne. Celem niniejszej pracy przeglądowej była prezentacja i omówienie wachlarza możliwości analitycznych dostępnych współczesnej biostatystyce i onkologii. Przegląd częstych problemów analitycznych i interpretacyjnych, możliwości projektowania i planowania badań oraz weryfikacji prawdziwości publikowanych danych może stanowić cenne uzupełnienie wiedzy onkologicznej, niezbędnej we współczesnej medycynie opartej na faktach i dowodach naukowych.

**Słowa kluczowe:** analiza przeżycia, hazard, modele prognostyczne

## ABSTRACT

Survival analysis is the analytical foundation of studies on cancer-related mortality or disease progression. Analytical techniques used for this purpose share one common trait — the uncertainty of the event's occurrence in individuals, whose observation time has been censored due to study termination, withdrawal due to events other than prespecified endpoints or loss to follow-up. The main advantage of such analytical methods is the possibility to estimate individual hazard (risk of event occurrence) at any given timepoint of observation and consequently of expected survival time depending on a plethora of clinical variables and treatment modalities. Due to a huge and ever-expanding amount of oncologic data using survival analysis as a primary measure of outcome, the ability to interpret such results and to know the assumptions and workings of particular methods is slowly becoming ubiquitous. Analytical methods deployed on survival data feature univariate nonparametric ones (the log-rank test), multivariate modeling techniques (assuming proportional or additive risk), automated neural networks and classification-regression trees. The purpose of this review was to present and discuss in detail the available range of analytic and exploratory methods

used in biostatistics and oncology. The spectrum of common analytical and interpretative problems, methods of designing and planning clinical trials and verifying the veracity of published data may provide a valuable addition in the process of clinical application of evidence based oncology.

**Key words:** survival analysis, hazard, prognostic models

Onkol. Prak. Klin. 2011; 7, 2: 89–101

## Wstęp

W zdecydowanej większości badań o tematyce onkologicznej najbardziej interesującym zagadnieniem jest to, czy badany schemat leczenia jest lepszy od standardowego lub placebo. Aby to ocenić, należy porównać między sobą badane grupy pod względem czasu do wystąpienia pewnego zdarzenia (którym najczęściej jest zgon, lecz może być to też innego rodzaju punkt końcowy). W przypadku zdarzeń, których wystąpienie nie budzi wątpliwości, a interwencja ma na celu usunięcie przyczyny choroby, kwestia definicji wyleczenia nie przedstawia szczególnych problemów. Przykładem może być dodatni posiew krwi potwierdzający obecność bakterii i terapia ukierunkowana na ich eradykację, której skuteczność potwierdza się wyjałowieniem kolejnego posiewu. W onkologii niepewność co do losów pacjenta jest znacznie większa — klinicysta nigdy nie ma pewności co do doszczętności wyleczenia choroby nowotworowej, a brak potwierdzonej wznowy może wynikać zarówno ze skutecznego leczenia, jak i zbyt krótkiego czasu obserwacji. Ta niepewność dotycząca losu pacjenta i ograniczone czasowo okno obserwacji wymuszają stosowanie specjalnych technik analizy statystycznej dopuszczających niepewność w zakresie wystąpienia efektu końcowego. W terminologii biostatystycznej techniki te nazywa się analizą przeżycia.

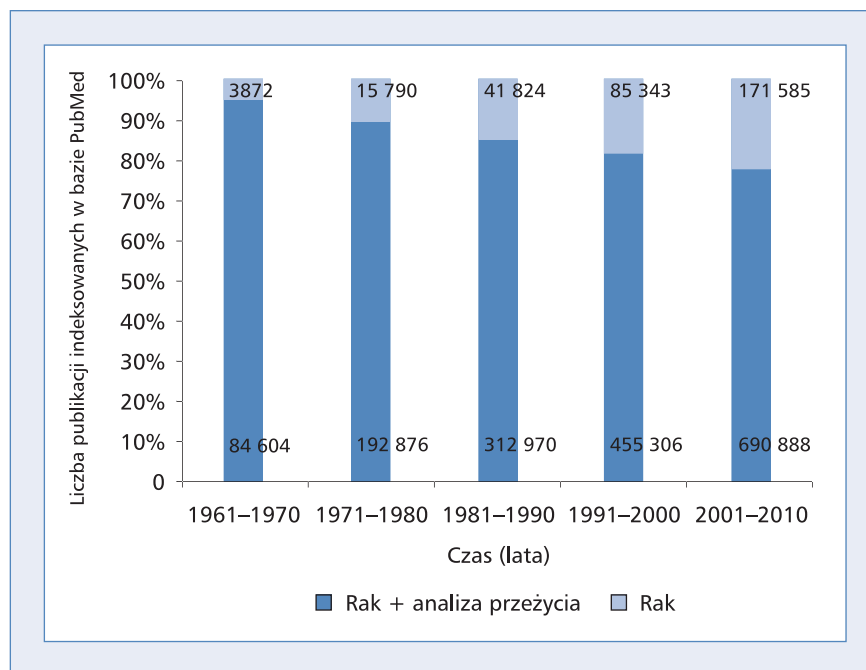
Analiza przeżycia (*survival analysis*) pozwala więc na wyznaczenie związku czynników badanych z punktem końcowym (czyli zdarzeniem będącym przedmiotem badań, np. zgonem, progresją choroby itp.) w sytuacjach, gdy informacja dotycząca wystąpienia punktu końcowego u wszystkich badanych jest ograniczona. Analizy tego typu nie są, jak mogłoby się wydawać, stosowane wyłącznie do analizy danych medycznych w kontekście przeżycia. W medycynie można je wykorzystywać do analizy w zasadzie dowolnego punktu końcowego występującego w czasie. Techniki analizy przeżycia stosuje się również między innymi w naukach społecznych i ekonomii jako analizę trwania (*duration analysis*), w przemyśle jako analizę niezawodności (*reliability analysis*) czy w demografii — jako analizę historii zdarzeń (*event history analysis*). We wszystkich tych przypadkach kluczową zaletą oferowaną przez opisane poniżej metody jest dopuszczalna niepewność co do dalszych losów obserwowanych pacjentów, zjawisk i zmiennych.

## Materiał i metody

Przeprowadzony przegląd piśmiennictwa na temat dostępnych metod analizy opierał się na głównej bazie danych indeksujących publikacje medyczne (PubMed). Dane wykorzystane do przedstawienia zależności biologicznych pochodziły z wcześniejszych prac własnych autorów. Wykorzystane słowa kluczowe obejmowały, między innymi, pojęcia z zakresu analizy przeżycia i pokrewnych problemów: „survival analysis”, „log-rank test”, „Cox proportional hazard model”, „survival modelling”, „Kaplan-Meier product limit metod” oraz terminologii onkologicznej: „cancer”, „neoplasm”, „oncology”, „malignancy”, „relapse”, „overall survival”, „progression-free survival” oraz „disease-free survival”. Przeprowadzone porównanie bibliometryczne liczby prac onkologicznych zawierających analizy przeżycia przeprowadzone u ludzi w ramach badań klinicznych obejmowało porównanie wyników wyszukiwania następujących kombinacji fraz: „cancer OR tumor OR neoplasm OR malignancy” dla prac wyłącznie onkologicznych oraz: „(cancer OR tumor OR neoplasm OR malignancy) AND („hazard ratio” OR „log-rank” OR „Kaplan-Meier” OR „survival analysis”)” dla grupy prac zawierających elementy analizy przeżycia. Ze względu na złożoność problemu, ogromną liczbę prac wykorzystujących techniki analizy przeżycia i planowany techniczno-edukacyjny charakter niniejszego opracowania nie przeprowadzono formalnej oceny listy publikacji zdalnych do przeprowadzenia działań typowych dla metaanalizy lub przeglądu systematycznego. Podane w dalszej części pracy przykłady, problemy i interpretacje wskaźników są więc ilustracją zjawiska o charakterze przede wszystkim edukacyjnym i kierującym działaniem medyczne lekarzy onkologów zainteresowanych stosowaniem założeń medycyny opartej na dowodach (EBM, *evidence based medicine*) w praktyce klinicznej.

## Wyniki

Przegląd piśmiennictwa, zgodnie z oczekiwaniami autorów, wykazał olbrzymią liczbę prac wykorzystujących analizy przeżycia lub skoncentrowanych na jej metodologii. Intrygująca była jednak nie tylko sama liczba prac z tej dziedziny, lecz również przyrost liczby



Rycina 1. Liczba prac z zakresu onkologii zawierających elementy analizy przeżycia zestawiona z liczbą prac onkologicznych opisujących inne efekty

Figure 1. The number of oncologic papers with application of survival analysis methods and those focused on other subjects over the last 50-years

prac onkologicznych wykorzystujących analizy przeżycia względem wszystkich prac z zakresu onkologii. Jakkolwiek analiza ta ze względu na specyficzne algorytmy wyszukiwania w bazie PubMed może być obciążona pewnym błędem, przedstawiona na rycinie 1 liczba prac z zakresu onkologii, w tym opierających się na analizach przeżycia, jest zdumiewająca.

Na przestrzeni ostatnich 50 lat liczba prac onkologicznych wykorzystujących analizy przeżycia zwiększyła się 50-krotnie, a udział tego typu publikacji wśród wszystkich prac onkologicznych wzrósł z 4% do 20%. Wszystko to działo się w sytuacji wykładniczego wzrostu liczby publikacji medycznych. Na podstawie tego spostrzeżenia można by wysnuć wniosek, że dla klinicystów nieposiadających umiejętności interpretacji tego typu wyników ponad 60% prac onkologicznych może być trudnych do zrozumienia. W XXI wieku — okresie, w którym zdecydowany nacisk kładzie się na wykorzystanie w praktyce klinicznej najnowszych doniesień naukowych — trudno jest praktykować świadomie EBM w onkologii, nie znając narzędzi analizy przeżycia. W dalszej części wyników zdefiniowano najważniejsze pojęcia z zakresu analizy przeżycia na podstawie kanonu piśmiennictwa z zakresu biostatystyki [1–4] oraz uzupełniono dodatkowymi informacjami na temat nowszych metod analizy lub bardziej szczegółowych problemów z ich stosowaniem.

## Podstawowe pojęcia w analizie przeżycia

Niepewność — jeśli punkt końcowy ma charakter definitywny i w chwili przeprowadzenia analizy wiadomo, że u żadnego z uczestników badania nie nastąpi zmiana jego stanu, można mówić o pewności co do punktu końcowego. Jeśli natomiast w chwili przeprowadzenia badania znany jest los jedynie osób osiągających punkt końcowy (zgon, progresja, wznowa potwierdzona badaniami), a w grupie porównawczej znany jest jedynie czas, przez jaki punkt końcowy nie wystąpił, można mówić o niepewności. W badaniach opierających się na punktach końcowych, w przypadku których jest całkowita pewność (przeżycie 30 dni po zabiegu operacyjnym, wypis z oddziału intensywnej terapii, zakończenie cyklu chemioterapii), nie stosuje się technik analizy przeżycia — w celu porównania czasu do wystąpienia takiego punktu końcowego wystarczający jest na przykład test U-Manna Whitneya.

## Czas przeżycia

Czas przeżycia definiuje się jako okres pomiędzy określonym punktem początkowym (tu np. zachorowaniem, rozpoczęciem leczenia itp.) a momentem wystąpienia punktu końcowego. Nie jest to zwykle tożsame z czasem trwania badania. Ponieważ przeważnie rekr-

tacja grupy badanej przebiega etapami, a pacjentów włącza się w kolejnych dniach czy miesiącach badania, u każdego z nich czas obserwacji jest indywidualny. Indywidualny czas przeżycia liczy się więc od chwili włączenia do badania do momentu wystąpienia punktu końcowego lub przerwania obserwacji pacjenta albo z powodu zakończenia badania czy przyczyn uniemożliwiających jego dalszą obserwację. Ponieważ każdego pacjenta obserwuje się przez różny okres i w chwili analizy wystąpienie punktu końcowego jest znane, czas ich przeżycia w badaniu onkologicznym można wyobrazić sobie tak, jak zilustrowano to na rycinach 2A i 2B.

### Obserwacje ucięte

Aby czerpać wiedzę z badań analizujących czas do wystąpienia punktu końcowego, którego nie obserwowano u większości badanych, należy przyjrzeć się nie tylko schematowi rekrutacji oraz kryteriom włączenia/wykluczenia z badania, ale także badanym punktom końcowym. Ważny jest też sposób, w jaki określano tzw. obserwacje ucięte (czyli obserwacje, dla których są pewne informacje o czasie przeżycia, ale dokładnie go nie określono). Obserwacje ucięte mogą wynikać z kilku różnych sytuacji klinicznych, takich jak:

- niewystąpienie u pacjenta (jeszcze) punktu końcowego (np. zgon, progresja nowotworu itd.) do momentu zakończenia badania;
- przerwanie obserwacji pacjenta (*lost to follow-up*) w okresie obejmującym badania;
- u pacjenta wystąpiła jakaś niepożądana reakcja, która uniemożliwiła jego dalszą obserwację do momentu ukończenia badania.

Powyższe sytuacje nie pozwalają na dokładne zmierzenie czasu do wystąpienia punktu końcowego, gdyż jest on poza zasięgiem terminu ukończenia badania. Taką sytuację nazywa się odcięciem do prawej strony.

Może zaistnieć również sytuacja, w której pacjenta obserwuje się w badaniu co pewien czas (np. co 3 miesiące) i w momencie kontroli pojawia się u niego punkt końcowy (np. wznowa kliniczna po zabiegu operacyjnym). Niestety nie wiadomo, kiedy dokładnie wystąpił moment punktu końcowego, mógł on pojawić się np. 1 miesiąc przed badaniem kontrolnym. Takie obserwacje ucięte nazywa się odcięciami do strony lewej.

Lepszym rozwiązaniem jest oczywiście ocena takich punktów końcowych, których brak wystąpienia skutkuje odcięciem do prawej strony, jednak są badania, w których stosuje się odcięcia do strony lewej. Często jest to konsekwencja przyjętego protokołu badawczego — w analizie przeżycia wolnego od progresji (PFS, *progression-free survival*) ocena czasu przeżycia zależy od wyniku badania obrazowego lub fizykalnego przeprowadzanego co pewien czas. Stwierdzenie progresji

choroby występujące pomiędzy jednym a drugim badaniem jest w istocie obserwacją uciętą do strony lewej. Do porównań PFS kluczowe jest więc założenie losowości rozkładu czasu do wystąpienia progresji pomiędzy badaniami w porównywanych grupach — jeśli w grupie kontrolnej badania będą opóźniane, czas do progresji zostaje sztucznie wydłużony. Zarówno pacjent, jak i lekarz nie znają przynależności do grupy co minimalizuje ryzyko oddziaływania obserwatora na czas przeżycia (*observation bias*), jednak nie zawsze jest to możliwe. W chorobach o wysokim prawdopodobieństwie wystąpienia punktu końcowego (hazardzie) efekt momentu oceny pacjenta widoczny jest na wykresach jako okresowo pojawiające się gwałtowne spadki skumulowanego prawdopodobieństwa przeżycia [5] (ryc. 3).

Jeśli efekt ten równoważy się pomiędzy grupami, nie będzie on oddziaływać na ocenę skuteczności leczenia, jakkolwiek dokładne oszacowanie oczekiwanego czasu przeżycia będzie obarczone dużym błędem. W przypadku gdy istnieją przesłanki pozwalające sądzić o tym, że w jednej grupie występowało znacznie więcej przypadków odstępstw od protokołu oceny lub też badanie nie było przeprowadzone z ślepą próbą, wyniki można uznać za mało wiarygodne. W związku z tym, że czas zgonu pacjenta jest zwykle znany z dokładnością co do dnia, efekt *observation bias* zasadniczo nie występuje w badaniach wykorzystujących analizę przeżycia całkowitego (OS, *overall survival*).

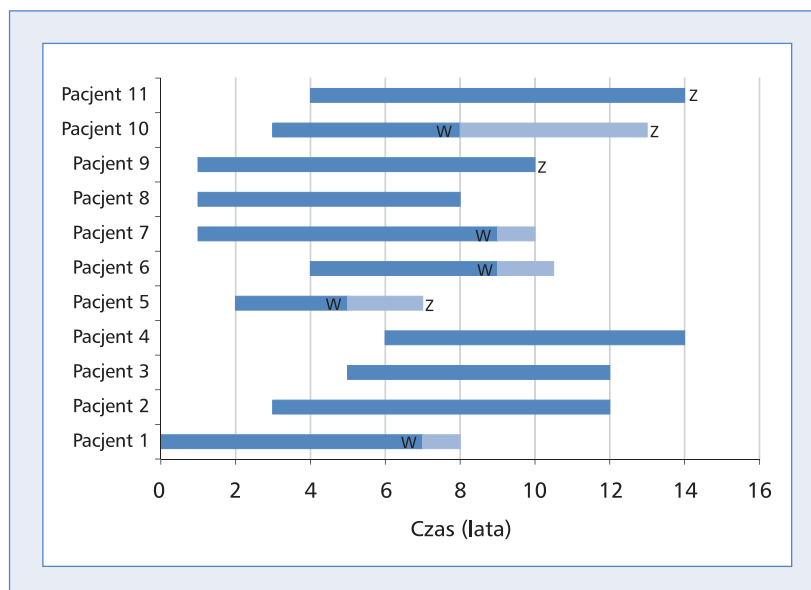
### Odcięcie informatywne

W sytuacjach, w których następuje przerwanie obserwacji pacjenta z określonego powodu z dużym prawdopodobieństwem związane z wystąpieniem punktu końcowego, można mówić o odcięciu informatywnym. Na przykład przerwanie obserwacji pacjenta w badaniu nad zaprzestaniem palenia przed zakończeniem programu edukacji lub farmakoterapii może być uznana za niepowodzenie leczenia [6]. W takich sytuacjach korzystne jest zdefiniowanie *a priori* sposobu, w jaki traktowane będą różnego rodzaju przyczyny przerwania obserwacji pacjenta. Jeśli informacje takie nie są zawarte w protokole badania, mogą sugerować próby „prania danych” i modyfikacji punktów końcowych po pierwszych analizach statystycznych wskazujących na brak zamierzonego przez badaczy efektu.

### Prawdopodobieństwo przeżycia i ryzyko zgonu

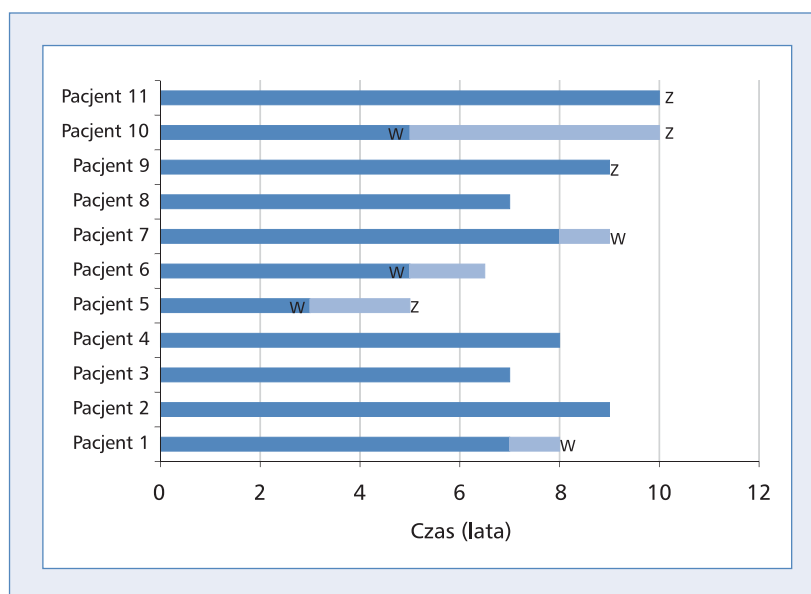
Dane dotyczące przeżycia opisuje się i wpisuje w modele statystyczne jako dwa ściśle ze sobą związane prawdopodobieństwa — przeżycia i zgonu.

Prawdopodobieństwo przeżycia (*survival probability*) jest prawdopodobieństwem przeżycia pojedynczej osoby w badaniu do końca czasu jego trwania. Jest to kluczo-



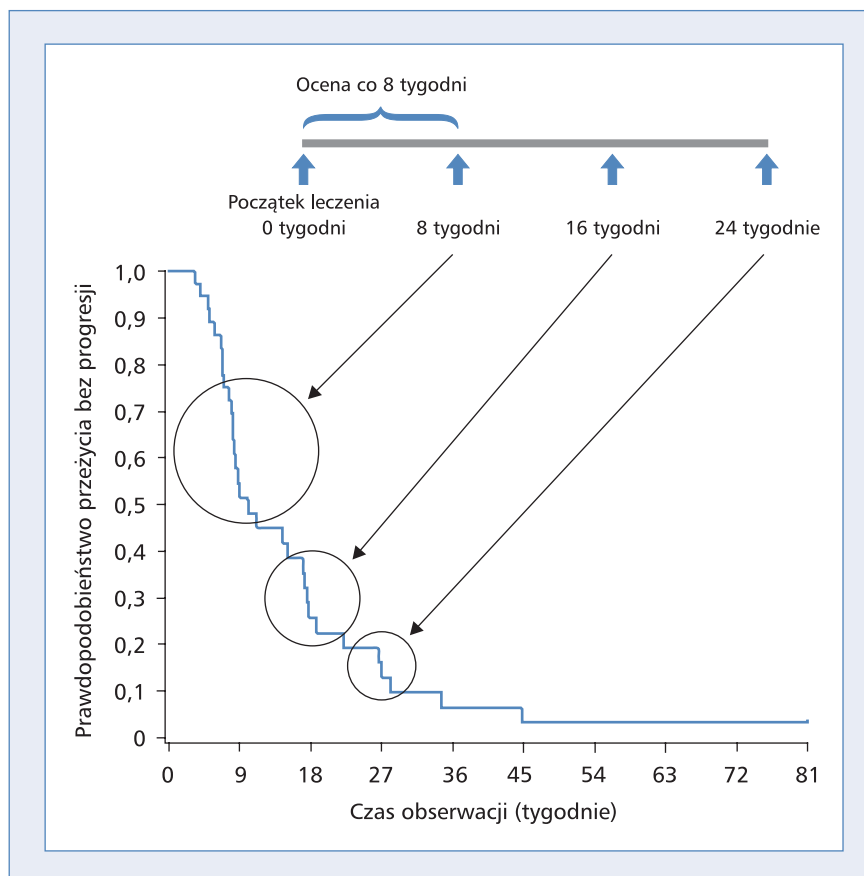
Rycina 2.A. Indywidualny czas przeżycia przykładowych pacjentów z uwzględnieniem różnego czasu włączenia do badania. Kolorem ciemnoszarym zaznaczono przeżycie po wznowie procesu nowotworowego. Wystąpienie punktu końcowego zaznaczono na indywidualnych liniach oznaczających pacjentów. W — wznowa procesu nowotworowego; Z — zgon

Figure 2.A. Variability of individual survival times of the studied patients due to different inclusion times. Dark grey areas show survival after relapse. The incidence of the analyzed endpoint was marked for respective patients: W — relapse; Z — patient's death



Rycina 2.B. Indywidualny czas przeżycia po wyrównaniu czasu obserwacji względem punktu początkowego. Kolorem ciemnoszarym zaznaczono przeżycie po wznowie procesu nowotworowego. Wystąpienie punktu końcowego zaznaczono na indywidualnych liniach oznaczających pacjentów. W — wznowa procesu nowotworowego; Z — zgon

Figure 2.B. Individual survival times of studied patients after normalization of starting date. Dark grey areas show survival after relapse. The incidence of the analyzed endpoint was marked for respective patients: W — relapse; Z — patient's death



Rycina 3. Efekt okresowo prowadzonej oceny progresji nowotworu w raku piersi leczonym bewacyzumabem — wyraźnie widać pojawiające się co 8 tygodni gwałtowne spadki ryzyka przeżycia ze względu na nieznaną dokładnie czas progresji i założenie analityczne, że wystąpił on w chwili oceny klinicznej

Figure 3. The effect of observation bias on cancer progression times in breast cancer undergoing bevacizumab treatment — evident losses of survival function occur every 8 weeks due to unknown exact progression time and the analytical bias entering the observation date as occurrence of the analyzed endpoint

we pojęcie w analizie przeżycia, ponieważ pojedyncze prawdopodobieństwa dla różnych czasów obserwacji dostarczają cennych informacji od momentu rekrutacji do zakończenia obserwacji. Następnie można dokonać podsumowania informacji otrzymanych z tych danych. Można je zobrazować w postaci wykresu (najczęściej mającego kształt funkcji schodkowej).

Ryzyko zgonu (hazard) jest to prawdopodobieństwo wystąpienia tego zdarzenia u konkretnego pacjenta włączonego do badania, żyjącego do końca czasu obserwacji. W kontekście wyłącznie funkcji ryzyka od czasu (nie w kontekście prawdopodobieństwa, ale funkcji matematycznej ryzyka zależnej od czasu  $t$ ) wartość hazardu jest chwilowym potencjałem dla wystąpienia badanego punktu końcowego. Gdy wartości funkcji przeżycia rosną, wartości funkcji ryzyka maleją, a ich suma pozostaje stała.

Ryzyko zgonu jest bardzo ważnym pojęciem w analizie przeżycia, gdyż na jego kanwie tworzy się modele przeżycia pozwalające na porównywanie efektu interwencji, ryzyka początkowego (stratyfikacji pacjentów na grupy prognostyczne) czy oceny potencjału terapeutycznego interwencji hipotetycznych znanych z badań przedklinicznych. Szacowanie efektu interwencji zwykle opiera się na mierze definiowanej jako współczynnik hazardu (HR, *hazard ratio*). Warto w tym miejscu zwrócić uwagę na lingwistyczną dwuznaczność pojęcia „ryzyko względne” (RR, *relative risk*) zarezerwowanego dla badań interwencyjnych, w których istnieje pewność co do punktu końcowego, oraz współczynnika hazardu tłumaczonego również jako współczynnik ryzyka, oznaczającego jednak funkcję ryzyka w czasie. Aby uniknąć nieporozumień terminologicznych, autorzy proponują jednak stosowanie wyrażenia współczynnik hazardu lub



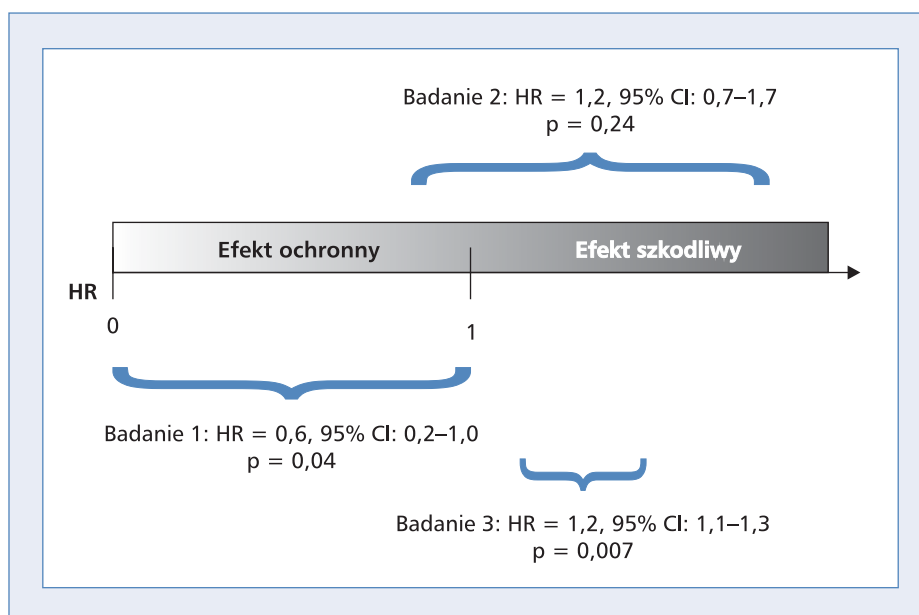
hazard względny, zamiast pojęć zawierających określenie „ryzyko” w badaniach obejmujących analizę przeżycia.

### Hazard względny

Hazard względny jest to stosunek ryzyka wystąpienia punktu końcowego w grupie A badania do ryzyka w grupie B w dowolnym punkcie czasowym. Wyliczany z ilorazu hazardu w grupie badanej względem porównawczej przyjmuje teoretycznie wartości od 0 (interwencja całkowicie zapobiega wystąpieniu punktu końcowego) do nieskończoności (interwencja zwiększa w największym możliwym stopniu prawdopodobieństwo wystąpienia punktu końcowego). Wartość 1 HR odpowiada interwencji idealnie neutralnej — zarówno w grupie badanej, jak i kontrolnej hazard punktu końcowego jest identyczny. Interpretacja współczynnika hazardu wymaga znajomości kilku dodatkowych matematycznych założeń. Miarą istotności statystycznej dla HR zwyczajowo jest 95-procentowy przedział ufności (95% CI, *confidence interval*). Wskazuje on z 95-procentowym prawdopodobieństwem zakres, w jakim dana wartość (w badaniach analizy przeżycia zwykle HR) znajduje w odniesieniu do całej populacji, a nie tylko grupy badanej. Przedział ufności pokazuje, jak dobrze oszacowano wartość miary

efektu (HR) w przeprowadzonym badaniu — w odniesieniu do populacji. Przedział ufności informuje więc o sile efektu, a zakres przyjmowanych wartości jest tożsamy z wartością istotności statystycznej ( $p$ ). Ponieważ wartość 1 oznacza interwencje idealnie neutralne, wartości mniejsze niż 1 oznaczają interwencje ochronne, a większe niż 1 szkodliwe, wynik istotny statystycznie będzie oznaczał taką interwencję, która z 95-procentowym prawdopodobieństwem przyniesie jednolity efekt. W praktyce oznacza to, że 95% CI, którego obie granice przedziału mieszczą się poniżej liczby 1, oznacza istotne statystycznie interwencje ochronne, a 95% CI, którego obie granice przedziału przekraczają wartość 1, oznacza interwencję o istotnej statystycznie szkodliwości. Na rycinie 4 zobrazowano przekrój możliwości wyników badań opisanych HR z 95% CI. Ponieważ 95% CI oblicza się na podstawie błędu standardowego HR, szerokość przedziału ufności zależy od liczebności grupy badanej. Interwencje nieistotne statystycznie (takie w których 95% CI zawiera liczbę 1) mogą nie oznaczać wyników bez znaczenia klinicznego, a jedynie dowodzić zbyt małej badanej grupy, na podstawie której badacze starali się udowodnić specyficzny efekt biologiczny.

Podsumowując, jeśli przykładowe HR liczone dla zgonu (OS) pacjentów stosujących lek A względem



Rycina 4. Przegląd przykładowych efektów badań. Badanie 1 obrazuje efekt ochronny, istotny statystycznie, o bardzo szerokim przedziale ufności. Badania 2 i 3 obrazują ten sam efekt, jednak ze względu na większą liczebność grupy 3 efekt tego badania osiągnął istotność statystyczną ( $p = 0,007$ ), podczas gdy w przypadku badania 2 odpowiadająca 95-procentowemu przedziałowi ufności (95% CI) wartość  $p$  wynosiła 0,24

Figure 4. Presentation of common study results. The study number 1 represents a protective effect with a wide 95% confidence interval (95% CI). Studies 2 and 3 represent the same effect size, but due to a great number of patients in study 3, its effect is statistically significant ( $p = 0.007$ ), while that of study 2 did not reach the threshold of significance with 95% CI values associated with a  $p$  value 0.24.

grupy kontrolnej wynosiło 0,76 z 95% CI: 0,6–0,92, oznaczało to, że lek A miał wpływ ochronny i zmniejszył prawdopodobieństwo zgonu o 24% w przebadanej grupie. W odniesieniu do całej populacji chorych na dany nowotwór oznacza to, że działanie leku z 95-procentowym prawdopodobieństwem będzie korzystne, a przedział HR będzie zawierał się między wartościami 0,60 i 0,92. Uzyskane w ten sposób ograniczenie prawdopodobieństwa zgonu o 8–40% mogłoby stanowić silną przesłankę do stosowania danego leku w praktyce klinicznej.

### Obrazowanie metod i analiza statystyczna czasu przeżycia — tablice trwania życia, metoda Kaplana-Meiera, testy i modele statystyczne. Na co zwrócić uwagę?

#### Tablice trwania życia

Najstarszym, lecz ciągle aktualnym sposobem opisywania czasów przeżycia są tablice trwania życia. Dzięki nim można najprościej opisać przeżycie w próbie. Narzędzie to wykorzystuje się w badaniu procesów migracyjnych i edukacyjnych, ale również w firmach zajmujących się ubezpieczeniami na życie.

Konstrukcja tablicy przeżycia opiera się na podziale czasu trwania obserwacji na przedziały klasowe o równej rozpiętości. Dla każdego z nich można obliczyć liczbę i proporcję pacjentów ciągle żyjących, poza obserwacją lub zmarłych. Opierając się na tych wartościach, oblicza się prawdopodobieństwo zgonu w jednostce czasu, liczbę chorych zagrożonych zgonem, a w efekcie funkcję przeżycia i hazardu. Mimo że tablice opisują przeżycie w próbie bardzo dobrze, to często potrzebna jest znajomość kształtu funkcji przeżycia w populacji. Wtedy pomocne okazują się różne modele dopasowania do danych 4 typowych rozkładów czasów przeżycia (wykładniczego, liniowego, Weibulla i Gompertza). Aby funkcję wpasować w dany model, należy sprawdzić zgodność tego dopasowania za pomocą testów statystycznych — zwykle przyrostowego testu  $\chi^2$  — weryfikując hipotezę zerową o braku różnicy modelu dopasowywanego i faktycznych danych. Podobnie jak w innych testach dopasowania rozkładu (W Shapiro-Wilka, Kołmogorowa-Smirnova), istotność statystyczna oznacza, że oba rozkłady (model i próba) różnią się, a populacyjne przeżycie z dużym prawdopodobieństwem rozkłada się w inny sposób niż w próbie badanej.

#### Mediana czasu przeżycia

Ponieważ czasy przeżycia w większości badań onkologicznych odbiegają znacznie od rozkładu normalnego,

w odniesieniu do czasu przeżycia zwyczajowo podaje się medianę zamiast średniej. Wartość mediany czasu przeżycia oznacza taki czas obserwacji, do którego u połowy obserwowanych pacjentów w danej grupie wystąpił punkt końcowy, a połowa pacjentów żyje nadal, wolna od tego samego punktu końcowego. Warto w tym miejscu nadmienić, że wszystkie metody statystyczne wymienione w dalszej części niniejszego opracowania nie porównują formalnie median czasu przeżycia pomiędzy grupami — wartość ta służy więc opisowi grup i oszacowaniu oczekiwanego czasu przeżycia, a nie obliczeniu istotności statystycznej.

#### Metoda Kaplana-Meiera

Tabele przeżycia i próby dopasowywania rozkładu są obarczone licznymi ograniczeniami, mianowicie koniecznością wymuszenia przedziałów czasu obserwacji o jednakowej długości. Analizę, która rozwiązuje ten problem, jest estymacja przeżycia metodą Kaplana-Meiera [7]. W technice tej nie ma konieczności tworzenia tabeli i dzielenia czasu na przedziały. Metoda wykorzystuje logiczne założenia, zgodnie z którymi chory, aby przeżyć 1 rok w badaniu, musi przeżyć wszystkie poprzednie okresy (miesiące po miesiącu) i i nie może u niego wystąpić punkt końcowy w 12. miesiącu indywidualnej obserwacji. Wzór wykorzystywany w tej analizie podano poniżej:

$$S(t_j) = S(t_{j-1}) \times \left(1 - \frac{d_j}{n_j}\right)$$

Z tego powodu, aby ocenić np. prawdopodobieństwo przeżycia w 12. miesiącu, mnoży się prawdopodobieństwo przeżycia kolejnych miesięcy, pod warunkiem przeżycia bez punktu końcowego (*event-free survival*) wszystkich poprzedzających. Takie prawdopodobieństwo nazywa się prawdopodobieństwem warunkowym. Krzywe powstałe na podstawie analizy Kaplana-Meiera są najczęściej stosowaną formą wizualizacji czasów przeżycia [8].

Podczas interpretacji krzywych Kaplana-Meiera warto zwracać uwagę na podpisy osi wykresów Y i X, gdyż idealnie separujące się krzywe mogą np. dotyczyć nie obserwacji miesięcy czy lat, a dni i oznaczać tym samym minimalną istotność kliniczną przy spektakularnej istotności statystycznej w teście log-rank. Jednocześnie oś Y może oznaczać zarówno OS, PFS, jak i prawdopodobieństwo innego rodzaju punktów końcowych, co wiąże się z diametralnie różną interpretacją efektu biologicznego stosowanej interwencji.

#### Weryfikacja hipotez statystycznych w analizie przeżycia

Aby ocenić wpływ pewnych zmiennych nominalnych (kategorii/grup) na czas przeżycia, funkcja przeżycia



musi zostać najpierw oszacowana dla obydwu grup. Następnie krzywe przedstawiające te funkcje porównuje się przy założeniu hipotezy zerowej o braku różnic prawdopodobieństwa zgonu pomiędzy grupami ( $HR = 1$ ). Na przestrzeni minionych lat powstało wiele narzędzi pozwalających na dokonywanie takich porównań. Statystyka posiada wiele narzędzi do weryfikacji takich hipotez, jednak najczęściej stosowanym w analizach przeżycia testem jest test logarytmiczny rang (test log-rank) [9, 10]. Test ten uważa się za standardowe narzędzie porównywania krzywych przeżycia, a wykorzystanie każdego innego w zasadzie uznaje się za niestandardowe [2]. Jeśli porównuje się jedynie dwie grupy, można obliczyć szacowane HR (wraz z możliwym do oszacowania 95% CI [11]) według wzoru:

$$HR = \frac{O_1/E_1}{O_2/E_2}$$

Gdzie  $O_1/E_1$  oznacza hazard w grupie 1, a  $O_2/E_2$  w grupie 2.

Test log-rank porównuje ze sobą 2 lub więcej krzywych przeżycia i bazuje na logarytmie funkcji przeżycia. Oblicza w każdym momencie czasu dla każdej grupy liczbę punktów końcowych, które wystąpiłyby, jeśli nie byłoby różnic pomiędzy grupami pod względem liczby pojawienia się tych punktów. Na koniec zlicza on liczbę tych punktów końcowych, które miałyby wystąpić, i porównuje te wartości z liczbą faktycznych punktów końcowych w każdej z grup pomiędzy badanymi krzywymi od czasu wystąpienia poprzedniego punktu końcowego. Test porównuje obserwowaną w grupach badanych liczbę punktów końcowych z hipotetyczną, przez siebie wyznaczoną wartością. Wartość testu log-rank porównuje się z rozkładem  $\chi^2$  o właściwej liczbie stopni swobody i na tej podstawie określa się wartość  $p$  — istotności statystycznej pomiędzy krzywymi. Jest to test nieparametryczny, czyli rozkład czasów do wystąpienia punktów końcowych nie musi spełniać założeń normalności rozkładu. Jednocześnie zmiana czasów przeżycia na rangi skutkuje dużą opornością testu log-rank na obserwacje odstające, co w analizie przeżycia odległego nowotworów o złym rokowaniu (rak płuca, rak trzustki) ma niebagatelne znaczenie — pacjentów z najdłuższym czasem przeżycia z czysto matematycznej perspektywy można byłoby uznać jako obserwacje odstające i usuwać z analizy parametrycznej.

Test log-rank uważa się za test dość mocny [12], czyli posiadający duży potencjał do odrzucenia hipotezy zerowej o równości hazardu w grupach. Należy jednak również pamiętać o jego ograniczeniach, z których głównym jest brak zakresu oszacowania efektu populacyjnego towarzyszącego wartości  $p$ .

Jak wspomniano wcześniej, istnieją również inne testy mające zastosowanie w analizie przeżycia, lecz

ich rozpowszechnienie jest zdecydowanie mniejsze niż testu log-rank: test Coxa-Mantela, test F Coxa (szczególnie polecany dla małych prób liczących poniżej 50 osób lub gdy zdecydowana większość obserwacji jest kompletna, a czas przeżycia nie odbiega istotnie od rozkładu wykładniczego lub Weibulla), test Wicoxona według Gehana czy test według Peto Peto. Ten ostatni stosuje się, gdy chce się zwrócić większą uwagę na początkowe części krzywych przeżycia. Ponieważ jednak ich zastosowania są ściślej określone, a założenia do ich wykorzystywania bardziej skomplikowane niż w przypadku testu log-rank, zamiar użycia tych testów w analizie przeżycia wymusza pytanie o metodologię i motywację autorów [2].

### Modele wieloczynnikowe

#### — model proporcjonalnego ryzyka Coxa

W analizach medycznych niemal zawsze więcej niż jedna zmienna determinuje hazard punktu końcowego. Jednocześnie wśród części zmiennych istnieje tendencja wzajemnych powiązań (np. zajęcie węzłów chłonnych i obecność przerzutów odległych). Aby rozdzielić analitycznie efekt poszczególnych czynników, wykorzystuje się techniki analizy wielowymiarowej zwane potocznie modelami. W analizie przeżycia najpowszechniej stosuje się półparametryczne modele proporcjonalnego ryzyka Coxa. W istocie są one modelem regresji wieloczynnikowej logarytmu hazardu, uwzględniającej efekt czasu i wynikającą z tego niepewność dotyczącą losów pacjentów. Formalny wzór modelu Coxa, w którym hazard  $[h(t)]$  szacuje się na podstawie  $p$  zmiennych ( $b_1, b_2, \dots, b_p$ ) podano poniżej:

$$h(t) = h_{0(t)} \times \exp \{b_1x_1 + b_2x_2 + \dots + b_px_p\}$$

Współczynniki  $b$  przypisane właściwym zmiennym oznaczają indywidualny wkład poszczególnych czynników do ostatecznego hazardu będącego wypadkową efektu obecności lub nieobecności poszczególnych składowych modelu. Wyraz wolny równania ( $h_0$ ) nazywa się hazardem podstawowym i odpowiada on prawdopodobieństwu wystąpienia punktu końcowego przy nieobecności wszystkich czynników uwzględnianych w modelu. Model Coxa zakłada, że hazard w porównywanych grupach pozostaje proporcjonalny na przestrzeni czasu obserwacji — w praktyce oznacza to, że odpowiadające funkcje Kaplana-Meiera nie mogą się przecinać. Analiza za pomocą modeli Coxa stanowi najczęściej stosowaną technikę wielowymiarowego modelowania przeżycia i pozwala na oszacowanie efektu zarówno zmiennych ciągłych, jak i nominalnych. Wieloskładnikowy wzór regresji Coxa pozwala również szacować zmianę hazardu punktu końcowego w zależności od modyfikacji poszczególnych czynników. Możliwe jest

więc szacowanie efektu zwiększenia dawki leku czy też hipotetyczny efekt włączenia leczenia u pacjentów w późniejszym/wcześniejszym stadium choroby. Zmienne uwzględnione w modelu Coxa nazywa się zmiennymi kowariancji (*covariates*), ponieważ ich efekt uwzględnia się łącznie i jednocześnie.

#### Modele wieloczynnikowe — brakujące dane zmiennych kowariancji

W większości medycznych baz danych pojawia się problem brakujących danych. O ile w przypadku analizy testem log-rank brak danej o przynależności do grupy jest raczej mało prawdopodobny, o tyle w sytuacji gdy analiza dotyczy 11 klinicznych zmiennych kowariancji, brak danych zaczyna stawać się problematyczny. Ponieważ pod względem formalnym brak danej na temat jednego składnika modelu Coxa powinien skutkować usunięciem pacjenta z analizy, konieczne jest opracowanie algorytmów postępowania w takich przypadkach. W tym celu stosuje się techniki wielokrotnego uzupełniania [13]. W ich przebiegu wprowadza się ustalone zestawy prawdopodobnych wartości, a wielokrotna analiza pozwala na ustalenie zakresu niepewności oszacowania wynikającego z uzupełnienia brakujących danych. O ile określa się je przed badaniem i wyjaśnia należycie w artykule, nie powinny wzbudzać szczególnych obaw — ostatecznie jeśli na przykład braki zostaną uzupełnione wartościami średnimi, a zmienna analizowana okaże się nieistotna statystycznie, nastąpi jej usunięcie z modelu i pozostawienie zmiennych, których wartości mogą być znane u wszystkich pacjentów. Jeśli więc algorytm postępowania z danymi brakującymi ustalono wcześniej i podano do wiadomości czytelników, obawy dotyczące manipulacji wynikiem są raczej bezpodstawne.

#### Modele wieloczynnikowe — selekcja zmiennych

Dostępność licznych zmiennych klinicznych powiązanych teoretycznie z hazardem punktu końcowego rodzi pokusę umieszczenia ich wszystkich w modelu wieloczynnikowym. W takich wypadkach jednak nie zawsze więcej znaczy lepiej. Część zmiennych może być pozbawiona znaczenia biologicznego w analizowanej grupie, część może pozostawać w silnej zależności od innych zmiennych, a jeszcze inne mogą mieć efekt wyrażony jedynie w niektórych specyficznych podgrupach (np. przy obecności mutacji specyficznego genu). Pojawia się więc problem selekcji zmiennych. Optymalne rozwiązanie w tej sytuacji nie istnieje. Pierwszą intuicyjną opcją jest włączanie zmiennych po kolei od najbardziej do najmniej istotnych, dopóki całość modelu pozostaje istotna (*forward selection*). Drugim prostym algoryt-

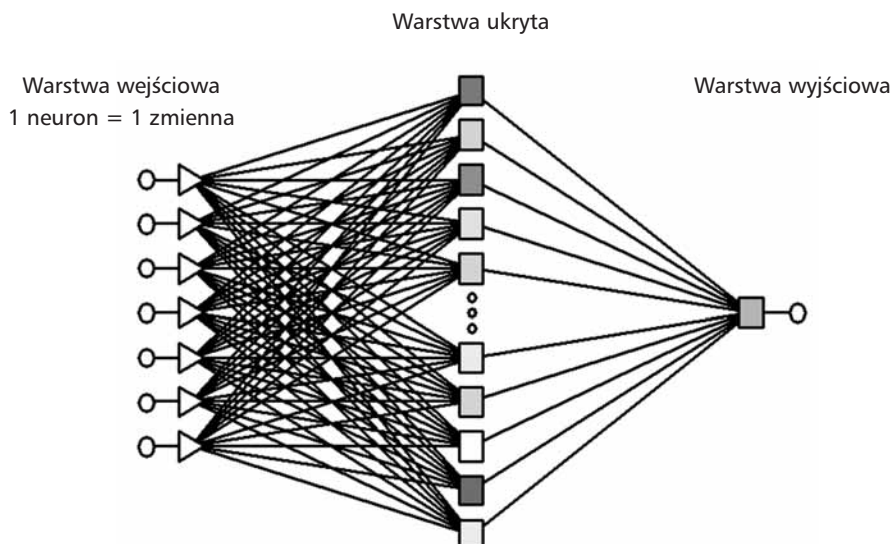
mem jest zbudowanie modelu opartego na wszystkich zmiennych i eliminacji najmniej istotnych (*backward elimination*). Obydwie metody nie gwarantują uzyskania najlepszego modelu, ponieważ obejmują jedynie niewielki wycinek całej przestrzeni możliwości, jednak alternatywne metody umożliwiające dogłębną eksplorację przestrzeni zmiennych są dużo bardziej czasochłonne, rzadko umieszczane w pakietach analizy statystycznej i w konsekwencji mało popularne [12]. Ogólnie przyjętym standardem jest metoda wstecznej selekcji zmiennych z sekwencyjnym wyłączeniem najmniej istotnych parametrów i każdorazową ewaluację modelu przed i po wyłączeniu zmiennej. Potocznie metodę tę nazywa się wsteczną, krokową selekcją zmiennych (*backward stepwise variable selection*).

### Inne techniki wieloczynnikowej analizy przeżycia

#### Sztuczne sieci neuronowe

Metodologia tworzenia sztucznych sieci neuronowych zakłada wykorzystanie samoadaptujących się neuronów (punktów decyzyjnych wykorzystujących funkcje matematyczne) do oszacowania finalnego efektu na podstawie danych wejściowych. Jakkolwiek metodologia wydaje się bardzo skomplikowana, efekt działania sieci neuronowej jest zwykle nieco lepszy niż klasycznych metod szacowania ryzyka lub równoważny [14–16]. Nie są one pozbawione jednak kilku zasadniczych wad utrudniających ich codzienne stosowanie i interpretację wyniku. W sieci neuronowej każda zmienna jest wprowadzana do modelu za pomocą neuronu pierwszej warstwy na kolejną warstwę ukrytą i wzbogacana o wagę (współczynnik). Każdy neuron wejściowy połączony jest z każdym neuronem warstwy ukrytej. Następnie każdy z neuronów warstwy ukrytej przekazuje otrzymane informacje o ważonych zmiennych na warstwę wyjściową, podającą zwykle prawdopodobieństwo punktu końcowego. Ponieważ w procesie uczenia sieci wagi dobierane są automatycznie, badacz i czytelnik nie mają żadnej możliwości określenia indywidualnego efektu poszczególnych zmiennych na przewidywany efekt interwencji. Z tego powodu sieci neuronowe określa się czasem jako „czarne skrzynki” — wewnętrzne działanie jest nieprzeniknione dla osoby oceniającej efekt ich prac. Schemat przykładowej sieci neuronowej podano na rycinie 5.

Sieci neuronowe wymagają również dużej liczby obserwacji, aby zwiększyć możliwości uogólnienia wyników predykcji — przy zbyt małej liczbie obserwacji sieć „uczy się na pamięć” podanych wartości i nie jest w stanie szacować efektu poza zakresem wartości zbioru treningowego.



Rycina 5. Schemat sieci neuronowej klasyfikacyjnej o typie perceptronu wielowarstwowego

Figure 5. A diagram of a neural net classifier based on a multilayer perceptron network

### Drzewa klasyfikacyjno-regresyjne

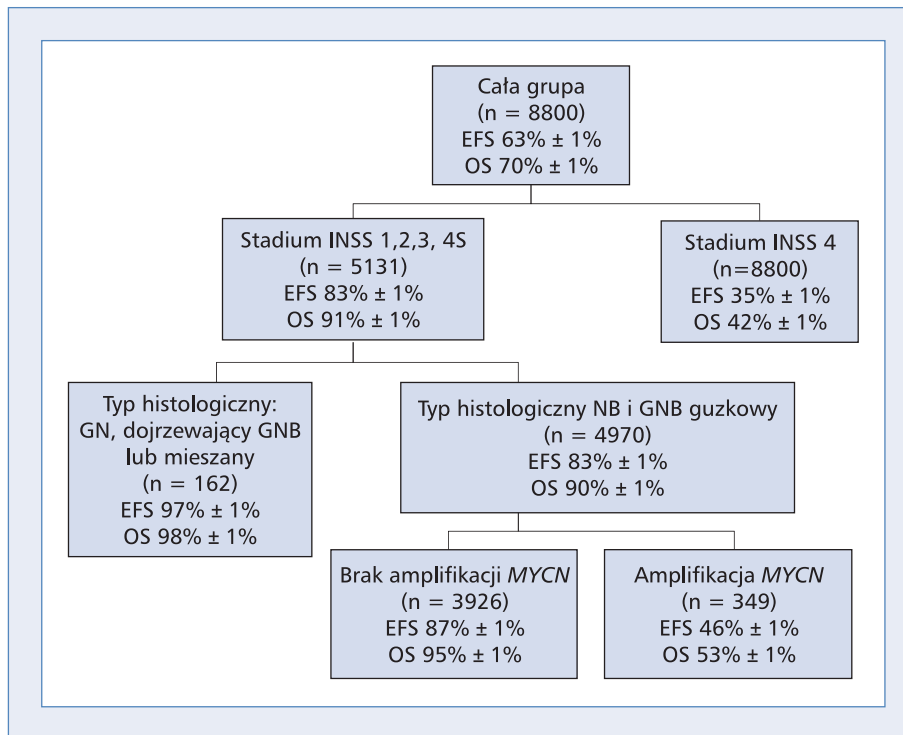
Niniejsze metody analityczne opierają się na dzieleniu grupy badanej na podgrupy na podstawie zmiennych o największym potencjale dyskryminacyjnym [17]. Wybór zmiennej dychotomizującej jest przeprowadzany na każdym kolejnym poziomie drzewa. W przypadku zmiennych ciągłych konieczne jest wybranie punktu odcięcia. Na rycinie 6 przedstawiono przykładowe drzewo klasyfikacyjne wykorzystane do prognozowania prawdopodobieństwa przeżycia pacjentów ze zwojakiem współczulnym zarodkowym [18].

Dzielenie podgrup kończy się w chwili, gdy żadna ze zmiennych nie przekracza ustalonego progu istotności statystycznej lub liczebność podgrup osiąga zamierzony próg (np. 25 osób). Zaletą drzew klasyfikacyjnych jest łatwość interpretacji i zastosowania praktycznego. Nie są one pozbawione wad — nadmierne dopasowanie do danych, brak uwzględnienia efektu czasu oraz konieczność arbitralnego ustalenia progu odcięcia zmiennych ciągłych utrudniają konstrukcję drzew klasyfikacyjnych we wszystkich sytuacjach klinicznych. Niemniej metoda ta idealnie nadaje się do prowadzenia badań eksploracyjnych lub projektowania schematów postępowania.

### Szacowanie liczebności grupy badanej potrzebnej do analizy przeżycia

Klasyyczna maksyma statystyczna „im więcej danych, tym lepiej” jest prawdziwa również w przypadku analizy przeżycia. Wiadomo jednak, że nie zawsze

możliwe jest zebranie grup liczących ponad kilkuset pacjentów, co może wynikać z rzadkości choroby, kosztów, czasu obserwacji lub czynników organizacyjnych. W takiej sytuacji stosuje się techniki estymacji pozwalające przewidzieć, jak liczna musi być grupa, aby z konkretnym prawdopodobieństwem (mocą statystyczną) badacz mógł odrzucić hipotezę zerową przy zadanym poziomie istotności statystycznej. Przeprowadzenie poprawnego szacowania wielkości grupy badanej wymaga w najbardziej podstawowej wersji: czasu rekrutacji, czasu obserwacji po zakończeniu rekrutacji, mediany czasu przeżycia w grupie kontrolnej (na tej podstawie oszacowuje się funkcję hazardu), planowanego efektu wyrażonego HR oraz pożądanej mocy statystycznej i poziomu istotności. Teoretycznie możliwe jest oszacowanie grupy wyłącznie na podstawie HR, jednak wymaga to spełnienia licznych warunków, w tym stałości hazardu w czasie (stałego odsetka pacjentów osiągających punkt końcowy w jednostce czasu), co w onkologii jest praktycznie niespotykane. Z tego powodu estymacja grupy wyłącznie na podstawie HR obarczona jest dużym błędem. W przypadku modeli wieloczynnikowych ogólna zasada sugeruje liczebność minimum 10–20 osób na grupę na każdą ze zmiennych kowariancji w modelu, przy czym ważniejsze są obserwacje, u których punkt końcowy wystąpił, niż obserwacje ucięte [1]. Samą procedurę konstrukcji optymalnego modelu i szacowanie hazardu najlepiej więc zostawić biostatystykom, jednak ogólna świadomość liczebności grup i informacji, o które można zostać zapytany podczas planowania badania, znacznie ułatwia przygotowanie protokołu badaczowi-klinicyście.



Rycina 6. Drzewo klasyfikacyjne przeżycia w zwojaku współczulnym zarodkowym. Na pierwszym etapie drzewa zmienną o największym potencjale dyskryminacji było zaawansowanie według skali INSS, na drugim — typ histologiczny guza, a na trzecim amplifikacja genu *MYCN*. EFS — przeżycie wolne od zdarzeń związanych z nowotworem; OS — przeżycie całkowite; GN — ganglioneuroma; GNB — ganglioneuroblastoma; INSS — *International Neuroblastoma Staging System*

Figure 6. A classification tree for survival in neuroblastoma. The first branching of the tree is dependent upon the most significant variable — INSS grade, the second dichotomization was made depending on histopathologic subtype and the third depending on *MYCN* gene amplification. EFS — event free survival; OS — overall survival; GN — ganglioneuroma; GNB — ganglioneuroblastoma; INSS — International Neuroblastoma Staging System

## Dyskusja

Celem niniejszego opracowania było przedstawienie wachlarza możliwości analiz przeżycia stosowanych powszechnie oraz błędów i problemów napotykanych podczas przeprowadzania, lektury lub stosowania praktycznych wyników badań onkologicznych. Przedstawione powyżej główne aspekty analizy przeżycia są oczywiście jedynie „wierzchołkiem góry lodowej” w tematyce technik analitycznych możliwych do zastosowania w analizie przeżycia. Niemniej jednak omówione zagadnienia obejmują większość najczęściej wykorzystywanych w pracach onkologicznych, a przykłady i wskazówki interpretacyjne mogą posłużyć za swego rodzaju instrukcję do świadomej lektury prac przedstawiających efekt czynników na przeżycie. Wśród bardziej złożonych technik analizy można by wymienić modele ryzyka addytywnego zakładające sumowanie się hazardu wynikającego z działania poszczególnych czynników [19]. W takich modelach czynnik biologiczny

może oddziaływać przez pewien czas jako ochronny, a w innych etapach badania jako neutralny lub nawet szkodliwy. Trudności interpretacyjne takich modeli [20] i wieloznaczność współczynników opisujących chwilowy efekt zmiennej na hazard podstawowy utrudniają stosowanie tych modeli w praktyce. Innym zagadnieniem z pogranicza analizy przeżycia i klasyfikacji jest problem kilku postępujących po sobie punktów końcowych. W rozważanych przykładach i większości literatury fachowej modele służą analizie do wystąpienia pierwszego punktu końcowego. Jeśli pewne punkty końcowe są od siebie uzależnione czasowo (progresja nowotworu i zgon) standardowy algorytm postępowania obejmuje przeprowadzenie odrębnych analiz poszczególnych punktów końcowych. Jeśli jednak takich analiz przeprowadza się wiele, badacze narażają się na zwiększone ryzyko potwierdzenia obserwacji fałszywie dodatnich. Alternatywą są modele warunkowe, do których pacjentów włącza się wielokrotnie, 1 raz dla każdego punktu końcowego, jaki wystąpił [21, 22].



Alternatywą są modele brzeżne, w których czasy do wystąpienia kolejnych punktów końcowych liczy się względem poprzednich. Złożoność analityczna tego typu metod przekracza ramy niniejszego opracowania, niemniej warto być świadomym możliwości obliczeniowych dostępnych statystykom.

Ze względu na powszechną dostępność oprogramowania statystycznego przeprowadzenie podstawowej analizy przeżycia z wykorzystaniem testu log-rank nie stanowi większego wyzwania dla zdeterminowanego badacza. Wszystkie poważne pakiety analityczne dostępne komercyjnie (SAS, STATISTICA, SPSS, Medcalc i inne) oraz darmowe narzędzia (R) oferują wiele rozmaitych metod analizy przeżycia, od najbardziej podstawowych do wyjątkowo rozbudowanych, wielopoziomowych modeli sekwencyjnych. Proste analizy zwykle skutkują uzyskaniem wyniku łatwego w interpretacji, jednak nie zawsze takie podejście jest najlepsze. Częstość bardziej skomplikowany model jest biologicznie i klinicznie wiarygodniejszy, a w konsekwencji bardziej prawdziwy, nawet za cenę trudności interpretacyjnej wyniku.

## Wnioski końcowe

Ze względu na olbrzymią liczbę prac onkologicznych pokazujących efekty badanych interwencji właśnie jako analizy przeżycia podstawowa znajomość założeń, technik, zalet i wad poszczególnych analiz jest koniecznym elementem arsenału każdego onkologa pragnącego leczyć pacjentów zgodnie ze standardami EBM. Wykorzystanie aktualnych dowodów naukowych, najnowszych wytycznych oraz wiedza o eksperymentalnych terapiach i projektach badawczych stają się bowiem standardem onkologii XXI wieku.

Wojciech Fendler i Wojciech Młynarski otrzymują wsparcie finansowe z programu *Polish Pediatric and Adolescent Diabetes Registry — Nationwide Genetic Screening for Monogenic Diabetes* finansowanego ze środków Programu Operacyjnego Innowacyjna Gospodarka I koordynowanego przez Fundację na Rzecz Nauki Polskiej.

## Piśmiennictwo

1. Bradburn M.J., Clark T.G., Love S.B., Altman D.G. Survival analysis Part III: multivariate data analysis — choosing a model and assessing its adequacy and fit. *Br. J. Cancer* 2003; 89: 605–611.
2. Clark T.G., Bradburn M.J., Love S.B., Altman D.G. Survival analysis part I: basic concepts and first analyses. *Br. J. Cancer* 2003; 89: 232–238.
3. Bradburn M.J., Clark T.G., Love S.B., Altman D.G. Survival analysis part II: multivariate data analysis — an introduction to concepts and methods. *Br. J. Cancer* 2003; 89: 431–436.
4. Clark T.G., Bradburn M.J., Love S.B., Altman D.G. Survival analysis part IV: further concepts and methods in survival analysis. *Br. J. Cancer* 2003; 89: 781–786.
5. Panageas K.S., Ben-Porat L., Dickler M.N., Chapman P.B., Schrag D. When you look matters: the effect of assessment schedule on progression-free survival. *J. Natl. Cancer Inst.* 2007; 99: 428–432.
6. Silagy C., Lancaster T., Stead L., Mant D., Fowler G. Nicotine replacement therapy for smoking cessation. *Cochrane Database Syst. Rev.* 2002; CD000146.
7. Kaplan E.L., Meier P. Nonparametric estimation from incomplete observations. *J. Am. Stat. Assoc.* 1958; 53: 457–481.
8. Bland J.M., Altman D.G. Survival probabilities the Kaplan-Meier method. *BMJ* 1998; 317: 1572.
9. Bland J.M., Altman D.G. The logrank test. *BMJ* 2004; 328: 1073.
10. Peto R., Pike M.C., Armitage P. i wsp. Design and analysis of randomized clinical trials requiring prolonged observation of each patient. II. analysis and examples. *Br. J. Cancer* 1977; 35: 1–39.
11. Collett D. Modelling survival data in medical research. Chapman & Hall, London 1994.
12. Hosmer D.W., Lemeshow S. Applied survival analysis: regression modelling of time to event data. Wiley, New York 1999.
13. van B.S., Boshuizen H.C., Knook D.L. Multiple imputation of missing blood pressure covariates in survival analysis. *Stat. Med.* 1999; 18: 681–694.
14. Sargent D.J. Comparison of artificial neural networks with other statistical approaches: results from medical data sets. *Cancer* 2001; 91: 1636–1642.
15. Linder R., König I.R., Weimar C., Diener H.C., Poppl S.J., Ziegler A. Two models for outcome prediction — a comparison of logistic regression and neural networks. *Methods Inf. Med.* 2006; 45: 536–540.
16. Snow P.B., Kerr D.J., Brandt J.M., Rodvold D.M. Neural network and regression predictions of 5-year survival after colon carcinoma treatment. *Cancer* 2001; 91: 1673–1678.
17. Lausen B., Sauerbrei W., Schumacher M. Classification and regression trees CART used for the exploration of prognostic factors measured in different scales. W: Dirschedl P., Osermann R. (red.). *Computational statistics*. Physica-Verlag, Heidelberg/New York 1994.
18. Cohn S.L., Pearson A.D., London W.B. i wsp. The International Neuroblastoma Risk Group INRG classification system: an INRG Task Force report. *J. Clin. Oncol.* 2009; 27: 289–297.
19. Wei L.J. The accelerated failure time model: a useful alternative to the Cox regression model in survival analysis. *Stat. Med.* 1992; 11: 1871–1879.
20. Aalen O.O. Further results on the non-parametric linear regression model in survival analysis. *Stat. Med.* 1993; 12: 1569–1588.
21. Tai B.C., Machin D., White I., GebSKI V. Competing risks analysis of patients with osteosarcoma: a comparison of four different approaches. *Stat. Med.* 2001; 20: 661–684.
22. Liu Y., Shete S., Etzel C.J. i wsp. Polymorphisms of LIG4, BTBD2, HMG2, and RTEL1 genes involved in the double-strand break repair pathway predict glioblastoma survival. *J. Clin. Oncol.* 2010; 28: 2467–2474.