

Piotr Król

Politechnika Rzeszowska

ESTYMACJA PODSTAWOWYCH FUNKCJI MODELU COXA – NA PODSTAWIE WYNIKÓW LECZENIA CHOROÓB NOWOTWOROWYCH

1. Wstęp

Istotną rolę w znacznej części badań medycznych, a zwłaszcza onkologicznych, odgrywa ocena czasu T przeżycia pacjentów poddawanych różnym metodom leczenia, z uwzględnieniem różnych, ale istotnych cech o charakterze endo- i egzogennym. We wspomnianych badaniach od momentu pokonania bariery obliczeniowej, coraz częściej wykorzystywany jest model Coxa, teoretycznie opracowany na przełomie lat siedemdziesiątych i osiemdziesiątych XX w. ubiegłego wieku [10]. W modelu tym ryzyko śmierci badanego pacjenta charakteryzuje iloczyn funkcji ryzyka bazowego zależnej od czasu i niezależnej od czasu eksponencjalnej funkcji liniowo uzależnionej od cech endo- i egzogennych.

W pracy rozpatrywany jest prosty model Coxa rozumiany w tym sensie, że w każdym momencie czasowym może zaistnieć co najwyżej jedno zdarzenie niepożądane¹ (np. zgon). Dla takiego modelu w literaturze [1; 8; 9]) opisuje się funkcję częściowej wiarygodności, umożliwiającą wyznaczenie wartości liczbowych parametrów niezbędnych w liniowej zależności cech endo- i egzogennych charakteryzujących pacjentów.

Dla funkcji ryzyka bazowego oraz dla bazowej funkcji przeżycia w zależności od przyjętych założeń i priorytetów wyprowadzone są różne estymatory o wysoce zróżnicowanym stopniu matematycznej złożoności. Jednym z trudniejszych [1; 5], ale, jak należy sądzić, najefektywniejszym, jest estymator Kalbfleischa–Prentice’a [4]. Dla modelu rozpatrywanego w pracy wyprowadzono analityczną postać wymienionego wyżej estymatora. Jednocześnie na przykładach umownych i rzeczy-

¹ Za zdarzenie niepożądane przyjmowany jest zwykle zgon jednostki kohorty, jakkolwiek mogą nim być inne zjawiska, np. nawrót choroby, wystąpienie objawów choroby itp. [1; 5; 7; 8; 9].

wistych pokazano, jak za pomocą tego estymatora można uzyskiwać przybliżone wartości funkcji ryzyka, funkcji przeżycia oraz współczynnika ryzyka.

2. Funkcja częściowej wiarygodności

Zbiorowość osobników ulegających różnym zdarzeniom i poddanych obserwacji nazywamy kohortą [8]. Liczebność kohorty $n = n(t)$ jest nierosnącą, prawostronnie ciągłą i przedziałami stałą funkcją czasu t . Na liczebność kohorty istotny wpływ mają badane zdarzenia niepożądane oraz inne zdarzenia, nazywane zdarzeniami cenzurowanymi². Jeśli $0 = t_0 < t_1 < t_2 < \dots < t_j < \dots < t_r$ oznaczają momenty czasowe wystąpienia zdarzenia niepożądanego na jednostkach kohorty, a $R_j = R(t_j)$ oznacza faktyczny (rzeczywisty) stan kohorty tuż przed zaistnieniem zdarzenia niepożądanego w chwili t_j , to $n_j - n_{j-1} \geq 1$ dla $j = 1, 2, \dots, r$, gdzie $n_j = n(t_j) = |R_j|$ oznacza liczebność zbioru R_j ($n_j - n_{j-1} > 1$ oznacza wystąpienie co najmniej jednego zdarzenia ocenzonego w przedziale czasu (t_{j-1}, t_j)).

Rozpatrywany w pracy model jest prostym modelem proporcjonalnej intensywności Coxa, co oznacza, że pojawienie się niepożądanego zdarzenia w momencie t_j jest możliwe tylko dla jednej jednostki. Jednostki kohorty nie muszą być identyczne. Różnicują je przeróżne cechy. Ze zbioru tych cech, kierując się przypuszczalnymi związkami przyczynowo-skutkowymi, wybieramy m cech oznaczonych przez $X_1, X_2, \dots, X_m$ lub w zapisie wektorowym $\mathbf{X} = [X_1, X_2, \dots, X_m]$. Współrzędne wektora wyróżnionych cech będziemy nazywać zmiennymi objaśniającymi³. Każdą więc j -tą jednostkę kohorty charakteryzuje układ m -wartości liczbowych zmiennych objaśniających $x_j = \langle x_{1j}, x_{2j}, \dots, x_{mj} \rangle$.

Zgodnie z ogólną teorią modelu proporcjonalnej intensywności Coxa [1; 3; 8; 9] funkcję ryzyka

$$h(\mathbf{x}; t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t},$$

czyli intensywności wystąpienia zdarzenia niepożądanego w momencie t na jednostce kohorty scharakteryzowanej wektorem $\mathbf{x} = \langle x_1, x_2, \dots, x_m \rangle$ definiuje się iloczynem

$$h(\mathbf{x}; t) = h_{\mathbf{x}}(t) := h_0(t) \exp(\boldsymbol{\beta} \cdot \mathbf{x}), \quad (1)$$

² Cenzurowanie to utrata jednostek kohorty z obserwacji bez odnotowania faktu zajścia zdarzenia niepożądanego. Zob. [7; 9; 10].

³ Zmienne objaśniające mogą mieć charakter ilościowy lub jakościowy. Realizacje zmiennych ilościowych wyrażane są w wartościach rzeczywistych a jakościowych poprzez odpowiednie rangowanie.

zależnej od czasu⁴ funkcji ryzyka bazowego $h_0(t)$ i eksponencjalnego czynnika uzależnionego tylko od iloczynu skalarnego $\beta \cdot x = \sum_{k=1}^m \beta_k \cdot x_k$, gdzie $x = \langle x_1, x_2, \dots, x_m \rangle$ jest wektorem cech charakteryzującym jednostkę, a $\beta = \langle \beta_1, \beta_2, \dots, \beta_m \rangle$ wektorem parametrów charakteryzującym całą kohortę. Stąd wzór

$$P(t \leq T < t + \Delta t | T \geq t) = h_0(t) \exp(\beta \cdot x) \Delta t + o(\Delta t), \quad (2)$$

gdzie $\lim_{\Delta t \rightarrow 0} \frac{o(\Delta t)}{\Delta t} = 0$, umożliwia analityczne przedstawienie funkcji częściowej wiarygodności $L(\beta)$ w zależności od parametrów $\beta = \langle \beta_1, \beta_2, \dots, \beta_m \rangle$:

$$L(\beta) = \prod_{j=1}^r P(B_j | A_j) = \prod_{j=1}^r \frac{\exp(\beta \cdot x_j)}{\sum_{i \in R_j} \exp(\beta \cdot x_i)} = \prod_{j=1}^r \frac{\exp \sum_{k=1}^m (\beta_k \cdot x_{kj})}{\sum_{i \in R_j} \exp \sum_{k=1}^m (\beta_k \cdot x_{ki})}, \quad (3)$$

gdzie B_j oznacza zdarzenie śmierci w chwili t_j jednostki scharakteryzowanej wektorem $x_j = \langle x_{1j}, x_{2j}, \dots, x_{mj} \rangle$, natomiast A_j zdarzenie zaistnienia tylko jednego zgonu w momencie t_j w zbiorze R_j [1].

Liczbowe wartości współrzędnych $\beta_1^*, \beta_2^*, \dots, \beta_m^*$ wektora parametrów β^* charakteryzującego całą kohortę i maksymalizującego funkcję częściowej wiarygodności, tzn. $L(\beta^*) = \text{MAX}$ są rozwiązaniami zadania optymalizacyjnego: $L(\beta) \rightarrow \text{MAX}$ (lub zadania z nim równoważnego $\ln L(\beta) \rightarrow \text{MAX}$).

3. Estymatory bazowych funkcji przeżycia i ryzyka

W procesie estymowania wartości funkcji ryzyka $h_x(t)$ oraz funkcji przeżycia⁵ $S_x(t)$ dla jednostki scharakteryzowanej wektorem cech $x = \langle x_1, x_2, \dots, x_m \rangle$ konieczne jest skonstruowanie estymatora $\hat{h}_0(t)$ funkcji $h_0(t)$ i estymatora $\hat{S}_0(t)$ funkcji $S_0(t)$ [1; 4; 9]. W literaturze przedmiotu [5; 8] rozpatrywane są różne estymatory tych funkcji wyprowadzone na bazie różnych przesłanek. Specyficzny charakter badań medycznych wymaga odmiennego podejścia. Dlatego też, celem pełnego zrozumienia poniższych przykładów, dokonano modyfikacji idei otrzymywania estymatorów $\hat{h}_0(t)$ i $\hat{S}_0(t)$.

⁴ $h_0(t) = h(0, t)$ – funkcja ryzyka dla umownej jednostki wzorcowej scharakteryzowanej zerowym wektorem cech $0 = \langle 0, 0, \dots, 0 \rangle$.

⁵ $S_x(t) = P(T \geq t)$, gdzie T jest zmienną losową opisującą czas życia, czyli czas do wystąpienia zdarzenia niepożądanego dla jednostki kohorty scharakteryzowanej wektorem x cech objaśniających.

W konstrukcji estymatorów bazowych funkcji ryzyka i przeżycia istotną rolę dla jednostki wzorcowej⁶ odgrywa liczbowe oszacowanie $\hat{h}_0(s)$ prawdopodobieństwa $P(t_{s-1} \leq T < t_s | T \geq t_{s-1})$ wystąpienia zdarzenia niepożądanego w lewostronnie domkniętych przedziałach czasowych, $[t_{s-1}, t_s)$ dla $s = 1, 2, \dots, j$ oraz $j = 1, 2, \dots, r$. Wtedy zależność $\xi(s) = 1 - \hat{h}_0(s)$ umożliwia oszacowanie prawdopodobieństwa:

$$1 - P(t_{s-1} \leq T < t_s | T \geq t_{s-1}) = P(T \geq t_s | T \geq t_{s-1}) = \frac{P(T \geq t_s)}{P(T \geq t_{s-1})},$$

przeżycia przedziału czasowego $[t_{s-1}, t_s)$ przez jednostkę scharakteryzowaną zerowym wektorem cech. Prawdopodobieństwo $S_0(t_j) = S_0(j)$ przeżycia przez jednostkę wzorcową przedziału czasowego $[0, t_j) = \bigcup_{s=1}^j [t_{s-1}, t_s)$ dla $j = 1, 2, \dots, r$, przy założeniu niezależności, jest iloczynem prawdopodobieństw przeżycia poszczególnych podprzedziałów $[t_{s-1}, t_s)$:

$$S_0(j) = \prod_{s=1}^j (1 - P(t_{s-1} \leq T < t_s | T \geq t_{s-1})) = \prod_{s=1}^j \frac{P(T \geq t_s)}{P(T \geq t_{s-1})} = P(T \geq t_j),$$

ponieważ $P(T \geq t_0 = 0) = 1$. Zatem za estymator $\hat{S}_0(j)$ bazowej funkcji przeżycia $S_0(j)$ przyjmujemy następujący iloczyn:

$$\hat{S}_0(j) = \prod_{s=1}^j \xi(s) = \prod_{s=1}^j (1 - \hat{h}_0(s)), \quad (4)$$

z którego w sposób oczywisty wynika równość:

$$\frac{\hat{S}_0(j)}{\hat{S}_0(j-1)} = \xi(j) \quad (5)$$

prawdziwa dla każdego $j = 1, 2, \dots, r$.

Dla jednostki wzorcowej mamy:

$$h(0, t) \Delta t = h_0(t) \Delta t = P(t \leq T < t + \Delta t | T \geq t) + o(\Delta t),$$

więc przy $t = t_{s-1}$, $\Delta t = t_s - t_{s-1}$ oraz $s = 1, 2, \dots, j$ iloraz

$$\frac{P(t_{s-1} \leq T < t_{s-1} + \Delta t | T \geq t_{s-1})}{t_s - t_{s-1}} = h_0(t_{s-1})$$

⁶ Jednostka wzorcową ($x = 0$) jest umowna i w rzeczywistości może w kohorcie nie występować.

określa przybliżoną wartość $h_0(t)$ w przedziale $[t_{s-1}, t_s)$. Ponieważ $\hat{h}_0(s)$ jest estymatorem prawdopodobieństwa $P(t_{s-1} \leq T < t_s | T \geq t_{s-1})$, więc jego relatywne odniesienie do długości przedziału czasowego $[t_{s-1}, t_s)$:

$$\hat{h}_0(s) = \frac{\hat{h}_0(s)}{t_s - t_{s-1}} = \frac{1 - \xi(s)}{t_s - t_{s-1}} \quad (6)$$

uznaje się za estymator funkcji ryzyka bazowego $h_0(t)$ dla $t \in [t_{s-1}, t_s)$.

Dla przykładu, gdy faktyczny stan kohorty $R(s)$ w przedziałach czasowych $[t_{s-1}, t_s)$ wynosi n_s jednostek jednakowo narażonych na zajście zdarzenia niepożądanego, to prawdopodobieństwo doznania niepowodzenia przez którąkolwiek z nich z nich wynosi $1/n_s$. Przyjmując $\xi(s) = 1 - 1/n_s$, $\hat{h}_0(s) = 1/n_s$, we wzorach (4) i (6) otrzymuje się szczególnie przypadek znanego w literaturze [6] estymatora Kaplana–Meiera dla bazowej funkcji przeżycia

$$\hat{S}_0(j) = \prod_{s=1}^j \left(1 - \frac{1}{n_s} \right) \quad (7)$$

oraz estymator dla bazowej funkcji ryzyka:

$$\hat{h}_0(j) = \frac{1}{n_j(t_{j+1} - t_j)}. \quad (8)$$

Estymatory (7) i (8) są niezależne od istotnych endo- i egzogennych cech charakteryzujących jednostki kohorty oraz od wektora β parametrów charakteryzujących całą kohortę. Tę wadę estymatorów Kaplana–Meiera eliminują estymatory Kalbfleischa–Prentice’a, których geneza dla prostego modelu proporcjonalnej intensywności jest następująca. Relatywne odniesienie prawdopodobieństwa $P(T \geq t_j)$ do prawdopodobieństwa $P(T \geq t_{j-1})$ na jednostce scharakteryzowanej wektorem $x = (x_1, x_2, \dots, x_m)$ jest równe:

$$\frac{P(T \geq t_j)}{P(T \geq t_{j-1})} = \frac{P(T \geq t_{j-1}) - P(t_{j-1} \leq T < t_j)}{P(T \geq t_{j-1})} = 1 - \frac{\exp \sum_{k=1}^m (\beta_k \cdot x_{kj})}{\sum_{l \in R_j} \exp \sum_{k=1}^m (\beta_k \cdot x_{kl})}.$$

Ze względu na to, że $S_x(t_j) = P(T \geq t_j)$ otrzymujemy:

$$\frac{S_x(t_j)}{S_x(t_{j-1})} = 1 - \frac{\exp \sum_{k=1}^m (\beta_k \cdot x_k)}{\sum_{l \in R_j} \exp \sum_{k=1}^m (\beta_k \cdot x_{kl})}. \quad (9)$$

Jednocześnie ze względu na równość $h_x(t_j) = h_0(t_j) \exp(\beta \cdot x)$ charakterystyczną dla modelu Coxa funkcję skumulowanego ryzyka⁷ wyrażamy następująco:

$$H_x(t_j) = H_0(t_j) \exp \sum_{k=1}^m \beta_k \cdot x_k,$$

gdzie $H_0(t_j) = \int_0^{t_j} h_0(\tau) d\tau$ jest bazowym ryzykiem skumulowanym. Funkcja przeżycia ma wtedy postać

$$S_x(t_j) = \exp(-H_0(t_j) \exp \sum_{k=1}^m \beta_k \cdot x_k) = [S_0(j)]^{\exp \sum_{k=1}^m \beta_k \cdot x_k},$$

$$\text{gdzie } S_0(j) = \exp[-H_0(t_j)] = \exp\left(-\int_0^{t_j} h_0(\tau) d\tau\right).$$

Stąd

$$\frac{S_x(t_j)}{S_x(t_{j-1})} = \left[\frac{S_0(j)}{S_0(j-1)} \right]^{\exp \sum_{k=1}^m \beta_k \cdot x_k},$$

zatem po uwzględnieniu (6):

$$\left[\frac{S_0(j)}{S_0(j-1)} \right]^{\exp \sum_{k=1}^m \beta_k \cdot x_k} = 1 - \frac{\exp \sum_{k=1}^m (\beta_k \cdot x_{kj})}{\sum_{l \in R_j} \exp \sum_{k=1}^m (\beta_k \cdot x_{kl})}. \quad (10)$$

Ponieważ t_j jest momentem zgonu jednej jednostki z kohorty $R(t_j)$, więc wzór (7) jest praktycznie użyteczny, gdy jest nią jednostka o cechach x . Dla zaakcentowania tego faktu przyjmujemy $x = x_j = \langle x_{1j}, x_{2j}, \dots, x_{mj} \rangle$. Wówczas, zamieniając w równości (7) funkcję S_0 jej estymatorem $\hat{S}_0$, otrzymujemy wzór analityczny umożliwiający wyznaczenie wartości estymatora $\xi(j)$ w przedziałach czasowych między dwoma kolejnymi zajściami zdarzeń niepożądanych z uwzględnieniem cech jednostek badanych:

⁷ W analizie przeżycia pomiędzy funkcjami: $h(t)$ – ryzyka, $f(t)$ – gęstości rozkładu prawdopodobieństwa losowego czasu życia T oraz $S(t)$ – przeżycia zachodzi związek $h(t)f(t) = S(t)$. Ryzyko skumulowane $H_x(t) = \int_0^t h_x(\tau) d\tau$ dla jednostki kohorty o cechach x jest powiązane z funkcją przeżycia $S_x(t)$ zależnością $H_x(t) = -\ln S_x(t)$ lub równoważnie $S_x(t) = \exp[-H_x(t)]$ (zob. np. [3; 9]).

$$\xi(j) = \left[1 - \frac{\exp \sum_{k=1}^m (\beta_k \cdot x_{kj})}{\sum_{l \in R_j} \exp \sum_{k=1}^m (\beta_k \cdot x_{kl})} \right]^{\exp(-\sum_{k=1}^m (\beta_k \cdot x_{kj}))} \quad (11)$$

Otrzymany wzór implikuje analityczną postać estymatora dla bazowej funkcji przeżycia, w literaturze [4] znanego jako estymator Kalbfleischa i Prentice'a:

$$\hat{S}_0(j) = \prod_{s=1}^j \xi(s) = \prod_{s=1}^j \left[1 - \frac{\exp \sum_{k=1}^m (\beta_k \cdot x_{ks})}{\sum_{l \in R_j} \exp \sum_{k=1}^m (\beta_k \cdot x_{kl})} \right]^{\exp(-\sum_{k=1}^m (\beta_k \cdot x_{ks}))}, \quad (12)$$

oraz analityczną postać estymatora funkcji ryzyka bazowego:

$$\hat{h}_0(s) = \frac{1}{t_s - t_{s-1}} \left[1 - \left[1 - \frac{\exp \sum_{k=1}^m (\beta_k \cdot x_{kj})}{\sum_{l \in R_j} \exp \sum_{k=1}^m (\beta_k \cdot x_{kl})} \right]^{\exp(-\sum_{k=1}^m (\beta_k \cdot x_{kj}))} \right], \quad (13)$$

dla $j = 1, 2, \dots, r-1$.

W otrzymanych estymatorach funkcji bazowych istotną rolę odgrywają parametry $\beta_1, \beta_2, \dots, \beta_m$ oraz cechy egzo- i endogenne $x = \langle x_1, x_2, \dots, x_m \rangle$ jednostki doznające zdarzenia niepożądanego oraz wszystkich pozostałych jednostek znajdujących się w badanej grupie tuż przed zaistnieniem kolejnych zdarzeń niepożądanych.

4. Relatywna ocena ryzyka i zasada proporcjonalności

W modelu Coxa wszystkie jednostki kohorty o tych samych $x = \langle x_1, x_2, \dots, x_m \rangle$ mają taką samą funkcję ryzyka $h(x, t)$. Ze wzoru (3) wiadomo, że jest ona równa iloczynowi funkcji ryzyka bazowego $h_0(t)$ oraz eksponenta współczynnika $\exp \beta \cdot x = \exp \left(\sum_{k=1}^m \beta_k \cdot x_k \right)$. Pełna wiedza o wartościach liczbowych parametrów $\beta_1, \beta_2, \dots, \beta_m$ charakteryzujących badaną kohortę i pełna wiedza o istotnych cechach endo- oraz egzogennych jakiejkolwiek jednostki nie wystarcza do jednoznacznego określenia jej funkcji ryzyka. W tym celu konieczna jest znajomość nie wyspecyfikowanej parametrycznie funkcji $h_0(t)$, którą można oszacować za pomocą estymatora (11) lub (16). Jednak znajomość $h_0(t)$ jest zbędna w przeprowadzaniu relatywnych ocen ryzyka dla różnych jednostek dożywających tego samego momentu czasowego $t, t > 0$. Pozwala na to założenie proporcjonalności w modelu Coxa, które stwierdza stały w czasie iloraz funkcji intensywności dla dowolnych dwóch jednostek grupy. Rzeczywiście, dla osobników z kohorty scharakteryzo-

wanych odpowiednio wektorami $x_j = \langle x_{1j}, x_{2j}, \dots, x_{mj} \rangle$ oraz $x_s = \langle x_{1s}, x_{2s}, \dots, x_{ms} \rangle$ zgodnie ze wzorem (3) jest:

$$HR_{j,s} = \frac{h_j(t)}{h_s(t)} = \frac{h_0(t) e^{\sum_{k=1}^m \beta_k x_{kj}}}{h_0(t) e^{\sum_{k=1}^m \beta_k x_{ks}}} = \exp \sum_{k=1}^m \beta_k (x_{kj} - x_{ks}). \quad (14)$$

Współczynnik $HR_{j,s}$ jest wielkością stałą w czasie dla jakichkolwiek dwóch jednostek scharakteryzowanych poprzez x_j i x_s ⁸. Iloraz $HR_{j,s}$ w literaturze [7; 9; 10] jest określany jako współczynnik ryzyka (ang. *hazard ratio*). Umożliwia on porównanie ryzyka wystąpienia zdarzenia niepożądanego dla dwóch różnych osobników z badanej grupy.

5. Wykorzystanie estymatorów Kalbfleischa-Prentice'a w przykładzie liczbowym

Przypuśćmy, że badana grupa liczy 6 pacjentów o czasach obserwacji $0 < t_1 < t_2 < \dots < t_l < \dots < t_r = 6$. Pacjenci obserwowani w momentach t_1, t_3, t_5, t_6 doświadczają zdarzenia niepożądanego (zgonu). Obserwacje pozostałych dwóch pacjentów w momentach t_2, t_4 , są cenzurowane⁹. Pacjentów różnicuje jedna zmienna objaśniająca $x = x$ przyjmująca wartość $x = 0$, gdy pacjent jest leczony, oraz $x = 1$, gdy pacjent nie jest leczony, a tylko obserwowany. Fakt cenzurowania l -tej obserwacji oznacza się przez $\delta_l = 0$, w przeciwieństwie do $\delta_l = 1$, gdy l -ta obserwacja nie podlega cenzurowaniu (jest kompletna). Funkcja intensywności zajścia zdarzenia niepożądanego dla modelu z tego przykładu jest postaci $h(t) = h_0(t) \exp(\beta x)$, gdzie $\beta x \in R$.

Dane z obserwacji badanej kohorty zawiera tab. 1.

Korzystając ze wzoru (3), można wyznaczyć funkcję częściowej wiarygodności $L(\beta)$. W badanej grupie występują tylko 4 obserwacje j zakończone zgonem, tzn. $j = 1$ dla $l = 1$, $j = 2$ dla $l = 3$, $j = 3$ dla $l = 5$, $j = 4$ dla $l = 6$. Zatem $L(\beta)$ jest iloczynem czterech czynników związanych z momentami czasowymi: $t_1 = 5$, $t_2 = 15$, $t_3 = 25$, $t_4 = 30$, z którymi związane są zbiory R_j , złożone z następujących pacjentów: $R_1 = \{1, 2, 3, 4, 5, 6\}$, $R_2 = \{3, 4, 5, 6\}$, $R_3 = \{5, 6\}$, $R_4 = \{6\}$. Każdy czynnik funkcji częściowej wiarygodności jest prawdopodobieństwem warunkowym, które oblicza się zgodnie ze wzorem (3).

⁸ Dla osobników o tych samych cechach ($x_j = x_s$) jest $HR_{j,s} = 1$.

⁹ Liczba porządkowa pacjenta jest utożsamiana z pacjentem (odpowiedniość wzajemnie jednoznaczna).

Tabela 1. Charakterystyka badanej grupy

Liczba porządkowa obserwowanych pacjentów	Czas obserwacji liczony w tygodniach	Wartość cechy	Wartość funkcji	Cenzurowanie	Liczba porządkowa zdarzeń niepożądanych	Zbiór pacjentów przed zaistnieniem niepowodzeń
l	t	x	e^{Ax}	δ	j	R_i
1	5	0	1	1	1	{1,2,3,4,5,6}
2	10	1	$\exp\beta$	0	–	–
3	15	1	$\exp\beta$	1	2	{3,4,5,6}
4	20	0	1	0	–	–
5	25	0	1	1	3	{5,6}
6	30	0	1	1	4	{6}

Źródło: dane umowne.

Dla pierwszego pacjenta zmarłego ($j = 1$ i $l = 1$):

$$P(B_1|A_1) = \frac{\exp(\beta \cdot x_{j=1})}{\sum_{l \in R_{j=1}} \exp(\beta \cdot x_l)} = \frac{\exp(0 \cdot \beta)}{\exp(0 \cdot \beta) + \exp(1 \cdot \beta) + \exp(1 \cdot \beta) + \exp(0 \cdot \beta) + \exp(0 \cdot \beta) + \exp(0 \cdot \beta)} =$$

$$= \frac{1}{4 + 2\exp\beta}$$

Dla drugiego pacjenta zmarłego ($j = 2$ i $l = 3$):

$$P(B_2|A_2) = \frac{\exp(\beta \cdot x_{j=2})}{\sum_{l \in R_{j=2}} \exp(\beta \cdot x_l)} = \frac{\exp(1 \cdot \beta)}{\exp(1 \cdot \beta) + \exp(0 \cdot \beta) + \exp(0 \cdot \beta) + \exp(0 \cdot \beta)} = \frac{\exp\beta}{3 + \exp\beta}$$

Dla trzeciego pacjenta zmarłego ($j = 3$ i $l = 5$):

$$P(B_3|A_3) = \frac{\exp(\beta \cdot x_{j=3})}{\sum_{l \in R_{j=3}} \exp(\beta \cdot x_l)} = \frac{\exp(0 \cdot \beta)}{\exp(0 \cdot \beta) + \exp(0 \cdot \beta)} = \frac{1}{2}.$$

Dla czwartego pacjenta zmarłego ($j = 4$ i $l = 6$):

$$P(B_4|A_4) = \frac{\exp(\beta \cdot x_{j=4})}{\sum_{l \in R_{j=4}} \exp(\beta \cdot x_l)} = \frac{\exp(0 \cdot \beta)}{\exp(0 \cdot \beta)} = 1.$$

Ostatecznie $L(\beta)$ ma postać:

$$L(\beta) = \prod_{j=1}^4 P(B_j|A_j) = P(B_1|A_1)P(B_2|A_2)P(B_3|A_3)P(B_4|A_4) = \frac{\exp\beta}{2(4 + 2\exp\beta)(3 + \exp\beta)},$$

stąd:

$$\ln L(\beta) = \beta - \ln(4 + 2e^\beta) - \ln(3 + e^\beta) - \ln 2.$$

Równanie

$$\frac{\partial [\ln L(\beta)]}{\partial \beta} = 1 - \frac{2e^\beta}{4+2e^\beta} - \frac{e^\beta}{3+e^\beta} = \frac{6-e^{2\beta}+2e^\beta}{(4+2e^\beta)(3+e^\beta)} = 0$$

jest równoważne z równaniem kwadratowym $6+2u-u^2=0$, w którym $e^\beta = u$. Równanie to ma dokładnie jedno rozwiązanie dodatnie¹⁰:

$$e^\beta = \frac{2+\sqrt{28}}{2} = 1+\sqrt{7} = 3,646.$$

Parametr charakteryzujący kohortę ma więc wartość $\beta = \ln(1+\sqrt{7}) \equiv \ln(3,646) \equiv 1,294$, a nie wyspecyfikowana funkcja ryzyka jest postaci: $h(t) = h_0(t) \exp(1,294x)$.

Oszacowana wartość liczbowa współczynnika $\beta = 1,294$ jest niewystarczająca do pozyskania pełnej postaci analitycznej funkcji ryzyka. Nieznany pozostaje człon $h_0(t)$ stanowiący zależne od czasu ryzyko bazowe. Natomiast wartość $\beta = 1,294$ jest wystarczająca do określenia stopnia siły oddziaływania zmiennej objaśniającej x na wystąpienie zdarzenia niepożądanego (zgonu). Miarą siły tego oddziaływania jest współczynnik ryzyka $HR_{j,s}$. Jeśli się przyjmie w przykładzie $x_j = 1$, $x_s = 0$, ze wzoru (14) wynika¹¹:

$$HR_{j,s} = \exp[\beta(x_j - x_s)] = \exp[1,294(1-0)] = \exp 1,294 = 3,646.$$

Współczynnik $HR_{j,s}$ określa, „ile razy” ryzyko wystąpienia zgonu dla pacjenta pierwszego (opisanego przez j -ty wariant zmiennej objaśniającej) jest większe w stosunku do tegoż ryzyka dla pacjenta drugiego (scharakteryzowanego przez s -ty wariant cechy). W analizowanym przykładzie ryzyko zgonu dla pacjenta obserwowanego ($x_j = 1$) jest 3,646-krotnie większe niż dla pacjenta leczonego ($x_s = 0$).

Dalsze analizy w badanej grupie będą możliwe pod warunkiem liczbowych oszacowań bazowej funkcji ryzyka $h_0(t)$ oraz bazowej funkcji przeżycia $S_0(t)$. Wartości liczbowe wspomnianych funkcji można szacować w oparciu o estymatory: $\hat{h}_0(t)$ i $\hat{S}_0(t)$ Kalbflleischa–Prentice’a przedstawione przez wzory (11), (12), (13). W tym celu całkowity czas obserwacji $[0, 30)$ należy rozbić na cztery podprzedziały: $[0, t_{j=1}) = [0, 5)$, $[t_{j=1}, t_{j=2}) = [5, 15)$, $[t_{j=2}, t_{j=3}) = [15, 25)$, $[t_{j=3}, t_{j=4}) = [25, 30)$ związane ze zgonami pacjenta pierwszego ($j = 1$), drugiego ($j = 2$), trzeciego ($j = 3$) i czwartego ($j = 4$). W każdym z nich wyznacza się zgodnie ze

¹⁰ Rozwiązanie ujemne $2 - \sqrt{28}/2$ odrzucamy, ponieważ z założenia: $e^\beta > 0$.

¹¹ Jeśli $x_j = x_j$, to $HR=1$. Jeśli $x_j = 0$ i $x_s = 1$, to $HR = 1/3,65 = 0,274$, natomiast w przypadku $x_j = x_s = 1$ (lub $x_j = x_s = 0$) jest $HR = 1$.

wzorem (11) wartość $\xi(j)$ estymatora przeżycia przez jednostkę wzorcową j -tego podprzedziału oraz wartość $\hat{h}_0(j)$ estymatora bazowej funkcji ryzyka dla $j = 1, 2, 3, 4$. W obliczeniach za wartość parametru charakteryzującego kohortę β przyjmuje się liczbę $\beta = 1,294$, którą otrzymano powyżej.

Dla $t \in [0,5)$ jest:

$$\begin{aligned}\xi(1) &= \left[1 - \frac{\exp \beta \cdot x_{j=1}}{\sum_{l \in R_{j=1}} \exp \beta \cdot x_l} \right]^{\exp(-\beta \cdot x_{j=1})} = \\ &= \left(1 - \frac{\exp(0 \cdot \beta)}{\exp(0 \cdot \beta) + \exp(1 \cdot \beta) + \exp(1 \cdot \beta) + \exp(0 \cdot \beta) + \exp(0 \cdot \beta) + \exp(0 \cdot \beta)} \right)^{\exp(-\beta \cdot 0)} = \\ &= \left(1 - \frac{1}{4 + 2e^\beta} \right)^{\exp(0)} = \left(\frac{3 + 2e^\beta}{4 + 2e^\beta} \right)^{\exp(0)} = \frac{3 + 2e^{1,294}}{4 + 2e^{1,294}} \cong 0,911\end{aligned}$$

$$\hat{h}_0(1) = 1 - \xi(1) = 1 - 0,911 \cong 0,089,$$

$$\hat{h}_0(1) = \frac{\hat{h}_0(1)}{t_j - t_{j-1}} = \frac{0,089}{5 - 0} = \frac{0,089}{5} \cong 0,018.$$

Dla $t \in [5,15)$ jest:

$$\begin{aligned}\xi(2) &= \left[1 - \frac{\exp \beta \cdot x_{j=2}}{\sum_{l \in R_{j=2}} \exp \beta \cdot x_l} \right]^{\exp(-\beta \cdot x_{j=2})} = \\ &= \left(1 - \frac{\exp(1 \cdot \beta)}{\exp(1 \cdot \beta) + \exp(0 \cdot \beta) + \exp(0 \cdot \beta) + \exp(0 \cdot \beta)} \right)^{\exp(-1 \cdot \beta)} = \\ &= \left(1 - \frac{e^\beta}{3 + e^\beta} \right)^{\exp(-1 \cdot \beta)} = \left(\frac{3}{3 + e^\beta} \right)^{\exp(-1 \cdot \beta)} = \left(\frac{3}{3 + e^{1,294}} \right)^{0,274} \cong 0,804\end{aligned}$$

$$\hat{h}_0(2) = 1 - \xi(2) = 1 - 0,804 \cong 0,196,$$

$$\hat{h}_0(2) = \frac{\hat{h}_0(2)}{t_j - t_{j-1}} = \frac{0,196}{15 - 5} = \frac{0,196}{10} \cong 0,020.$$

Dla $t \in [15,25)$ jest:

$$\begin{aligned}\xi(3) &= \left[1 - \frac{\exp \beta \cdot x_{j=3}}{\sum_{l \in R_{j=3}} \exp \beta \cdot x_l} \right]^{\exp(-\beta \cdot x_{j=3})} = \left(1 - \frac{\exp(0 \cdot \beta)}{\exp(0 \cdot \beta) + \exp(0 \cdot \beta)} \right)^{\exp(-\beta \cdot 0)} = \\ &= \left(1 - \frac{1}{2} \right)^{\exp(0)} = \left(\frac{1}{2} \right) = 0,500\end{aligned}$$

$$\hat{h}_0(3) = 1 - \xi(3) = 1 - \left(\frac{1}{2} \right) = 0,500,$$

$$\hat{h}_0(3) = \frac{\hat{h}_0(3)}{t_j - t_{j-1}} = \frac{0,500}{25 - 15} = \frac{0,500}{10} = 0,050$$

Dla $t \in [25,30)$ jest:

$$\xi(4) = \left[1 - \frac{\exp \beta \cdot x_{j=4}}{\sum_{l \in R_{j=4}} \exp \beta \cdot x_l} \right]^{\exp(-\beta \cdot x_{j=4})} = \left(1 - \frac{\exp(0 \cdot \beta)}{\exp(0 \cdot \beta)} \right)^{\exp(-\beta \cdot 0)} = (1 - 1)^{\exp(0)} = 0,$$

$$\hat{h}_0(4) = 1 - \xi(4) = 1 - (0) = 1,$$

$$\hat{h}_0(4) = \frac{\hat{h}_0(4)}{t_j - t_{j-1}} = \frac{1}{30 - 25} = \frac{1}{5} = 0,2.$$

Bazowa funkcja przeżycia $S_0(t)$ estymowana jest funkcją $\hat{S}_0(t)$, której wartości są obliczone zgodnie ze wzorem (12). Poniżej przedstawiono zestawienie obliczeń i wartości estymatora $\hat{S}_0(t)$ w poszczególnych podprzedziałach, dla $\beta = 1,294$:

$$\hat{S}_0(t) = \begin{cases} \left(\frac{3 + 2e^\beta}{4 + 2e^\beta} \right)^{\exp(0)} = 0,911 \text{ dla } t \in [0,5) \\ \left(\frac{3 + 2e^\beta}{4 + 2e^\beta} \right)^{\exp(0)} \cdot \left(\frac{3}{3 + e^\beta} \right)^{\exp(-\beta)} = 0,911 \cdot 0,804 = 0,732 \text{ dla } t \in [5,15) \\ \left(\frac{3 + 2e^\beta}{4 + 2e^\beta} \right)^{\exp(0)} \cdot \left(\frac{3}{3 + e^\beta} \right)^{\exp(-\beta)} \cdot \left(\frac{1}{2} \right)^{\exp(0)} = 0,911 \cdot 0,804 \cdot 0,5 = 0,366 \text{ dla } t \in [15,25) \\ \left(\frac{3 + 2e^\beta}{4 + 2e^\beta} \right)^{\exp(0)} \cdot \left(\frac{3}{3 + e^\beta} \right)^{\exp(-\beta)} \cdot \left(\frac{1}{2} \right)^{\exp(0)} \cdot (0) = 0,911 \cdot 0,804 \cdot 0,5 \cdot 0 = 0 \text{ dla } t \in [25,30) \end{cases}$$

Otrzymane estymatory $\hat{S}_0(t)$ i $\hat{h}_0(t)$ pozwalają obliczyć dla dowolnej jednostki kohorty scharakteryzowanej przez wektor cech objaśniających x ich indywidualne

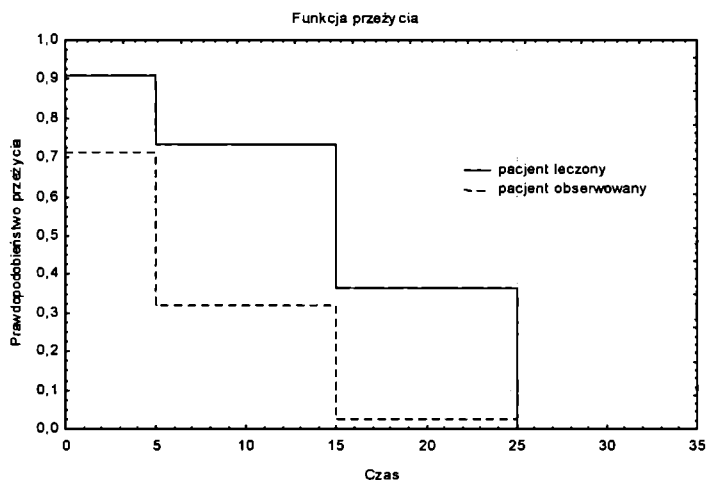
funkcje przeżycia $S_x(t_j) \equiv \hat{S}_0(j)^{\exp \beta x}$ i ryzyka (intensywności zaistnienia zdarzenia niepożądanego): $h_x(t) = h(x; t) = \hat{h}_0(t) \exp(\beta \cdot x)$.

Tabela 2. Funkcje ryzyka i przeżycia dla osobników badanej grupy

Przedział czasowy $[t_{j-1}, t_j)$	Osobnik $x = 0$		Osobnik $x = 1$	
	Funkcja ryzyka $h(x=0; t)$	Funkcja przeżycia $S_{x=0}(t)$	Funkcja ryzyka $h(x=1; t)$	Funkcja przeżycia $S_{x=1}(t)$
[0,5)	0,018	0,911	0,066	0,712
[5,15)	0,02	0,732	0,073	0,321
[15,25)	0,05	0,366	0,182	0,026
[25,30)	0,2	0	0,729	0

Źródło: opracowanie własne.

Badane funkcje przeżycia przedstawiono także na rys. 1.



Rys. 1. Wykresy funkcji przeżycia dla pacjentów leczonych i obserwowanych

Źródło: opracowanie własne.

W przykładzie występuje tylko jedna dychotomiczna zmienna objaśniająca (przybierająca wartości 0 oraz 1). Dla pacjenta leczonego, czyli scharakteryzowanego zmienną $x = 0$ funkcje przeżycia i ryzyka są tożsame z funkcjami bazowymi, ponieważ $S_{x=0}(t_j) \equiv [\hat{S}_0(j)]^{\exp(0)} = S_0(j)$ oraz $h(x=0; t) \equiv \hat{h}_0(t) \exp(0) = \hat{h}_0(t)$. Natomiast dla pacjenta nie leczonego mającego wartość cechy $x=1$, funkcja przeżycia $S_{x=1}(t_j)$ przedziału czasu $[0, t_j)$ jest estymowana funkcją $[\hat{S}_0(j)]^{3,646}$,

a funkcja ryzyka dla tej jednostki $h(x=1; t_j)$ jest estymowana funkcją $\hat{h}_0(t_j)e^{1,294} = 3,646\hat{h}_0(t_j)$. Ostatecznie, wyestymowane wartości funkcji ryzyka oraz przeżycia dla obu osobników przedstawiono w tab. 2.

6. Wykorzystanie estymatorów Kalbfleischa–Prentice’a w przykładzie liczbowym – wariant z dwoma zmiennymi objaśniającymi

W zależności od zaawansowania choroby nowotworowej (miejscowe – guz pierwotny vs. regionalne – guz i zajęte okoliczne węzły chłonne) porównywano dwie metody leczenia: A i B. Oceniano czas przeżycia t liczony od momentu wdrożenia leczenia do zajścia zdarzenia niepożądanego – zgonu. Obserwacje dla czterech pacjentów są niekompletne (cenzurowane), tzn. $\delta = 0$. Pacjentów badanej kohorty scharakteryzowano przez dwie zmienne objaśniające: X_1 – metoda leczenia, X_2 – zaawansowanie choroby. Chorych leczonych metodą A oznaczono jako $x_1 = 1$, metodą B jako $x_1 = 0$. Pacjentów z chorobą o zaawansowaniu miejscowym opisano jako $x_2 = 1$, a w zaawansowaniu regionalnym jako $x_2 = 0$. Badana grupa liczy 14 pacjentów, których charakterystyki zawarto w poniższej tab. 3.

Tabela 3. Charakterystyka badanej grupy

Liczba porządkowa obserwowanych pacjentów	Czas przeżycia liczony w tygodniach	Cenzurowanie	Liczba porządkowa zdarzeń niepożądanych	Metoda leczenia	Zaawansowanie choroby
l	T	δ	j	x_1	x_2
1	2	1	1	0	0
2	5	1	2	0	1
3	8	0	–	1	0
4	10	1	3	0	0
5	12	0	–	1	1
6	15	1	4	1	0
7	20	1	5	0	0
8	22	0	–	0	1
9	25	1	6	0	0
10	28	1	7	1	0
11	30	0	–	1	1
12	35	1	8	1	1
13	40	1	9	0	1
14	50	1	10	1	1

Źródło: dane umowne.

Korzystając ze wzoru (3), można wyznaczyć funkcję częściowej wiarygodności $L(\beta)$. Idea utworzenia tej funkcji jest identyczna jak w przykładach powyższych. Jest to iloczyn 10 prawdopodobieństw warunkowych związanych ze zgonami poszczególnych osobników badanej grupy.

Dla pierwszego ($l = 1, j = 1$) pacjenta jest:

$$P(B_1|A_1) = \frac{1}{4 + 3e^{\beta_1} + 3e^{\beta_2} + 4e^{(\beta_1+\beta_2)}}.$$

Dla drugiego ($l = 2, j = 2$) pacjenta jest:

$$P(B_2|A_2) = \frac{e^{\beta_2}}{3 + 3e^{\beta_1} + 3e^{\beta_2} + 4e^{(\beta_1+\beta_2)}}.$$

Dla czwartego ($l = 4, j = 3$) pacjenta jest:

$$P(B_3|A_3) = \frac{1}{3 + 2e^{\beta_1} + 2e^{\beta_2} + 4e^{(\beta_1+\beta_2)}}.$$

Dla szóstego ($l = 6, j = 4$) pacjenta jest:

$$P(B_4|A_4) = \frac{e^{\beta_1}}{2 + 2e^{\beta_1} + 2e^{\beta_2} + 3e^{(\beta_1+\beta_2)}}.$$

Dla siódmego ($l = 7, j = 5$) pacjenta jest:

$$P(B_5|A_5) = \frac{1}{2 + e^{\beta_1} + 2e^{\beta_2} + 3e^{(\beta_1+\beta_2)}}.$$

Dla dziewiątego ($l = 9, j = 6$) pacjenta jest:

$$P(B_6|A_6) = \frac{1}{1 + e^{\beta_1} + e^{\beta_2} + 3e^{(\beta_1+\beta_2)}}.$$

Dla dziesiątego ($l = 10, j = 7$) pacjenta jest:

$$P(B_7|A_7) = \frac{e^{\beta_1}}{e^{\beta_1} + e^{\beta_2} + 3e^{(\beta_1+\beta_2)}}.$$

Dla dwunastego ($l = 12, j = 8$) pacjenta jest:

$$P(B_8|A_8) = \frac{e^{\beta_1} \cdot e^{\beta_2}}{e^{\beta_2} + 2e^{(\beta_1+\beta_2)}}.$$

Dla trzynastego ($l = 13, j = 9$) pacjenta jest:

$$P(B_9|A_9) = \frac{e^{\beta_2}}{e^{\beta_2} + e^{(\beta_1+\beta_2)}}.$$

Dla czternastego ($l = 14, j = 10$) pacjenta jest:

$$P(B_{10}|A_{10}) = \frac{e^{(\beta_1 + \beta_2)}}{e^{(\beta_1 + \beta_2)}} = 1.$$

Funkcja częściowej wiarygodności $L(\beta)$ przyjmuje postać wielomianu wysokiego stopnia, którego miejsca zerowe wyznaczono z użyciem pakietu Statistica 6.0 PL. Otrzymano następujące wartości współczynników regresji: $\beta_1 = -0,997$ oraz $\beta_2 = -2,275$.

Oszacowane wartości parametrów β_1, β_2 pozwalają wyliczyć, zgodnie ze wzorem (14) współczynniki ryzyka HR dla obu zmiennych objaśniających. Jeśli $x_{2j} = x_{2s}$ oraz jeśli $x_{1j} = 1$ i $x_{1s} = 0$, to $HR = 0,36896$ i odwrotnie, gdy $x_{1j} = 0$ i $x_{1s} = 1$, to $HR = 2,710321$. Natomiast, gdy $x_{1j} = x_{1s}$ oraz jeśli $x_{2j} = 1$ i $x_{2s} = 0$ to $HR = 0,102764$ oraz analogicznie dla $x_{2j} = 0$ i $x_{2s} = 1$ $HR = 9,731013$. Przypadki, gdy pacjenci różnią się między sobą obydwooma cechami, przedstawiono w tab. 4.

Tabela 4. Współczynniki ryzyka dla przykładowych pacjentów

Pacjent	Wartość x_1	Wartość x_2	Wartość HR
<i>j</i> -ty	1	0	3,5903
<i>s</i> -ty	0	1	
<i>j</i> -ty	1	1	0,0379
<i>s</i> -ty	0	0	
<i>j</i> -ty	0	0	26,3742
<i>s</i> -ty	1	1	
<i>j</i> -ty	0	1	0,2785
<i>s</i> -ty	1	0	

Źródło: opracowanie własne.

Analiza współczynników ryzyka wykazuje, że przy identycznej wartości zmiennej x_2 zagrożenie wystąpieniem zdarzenia niepożądanego jest 2,71-krotnie większe dla pacjenta leczonego metodą B w stosunku do chorego leczonego metodą A. Ryzyko zgonu jest również 9,73-krotnie wyższe dla osobników z chorobą o zaawansowaniu regionalnym w stosunku do tych o miejscowej lokalizacji nowotworu, o ile byli leczeni tą samą metodą.

Dalsze analizy o konkretnych przebiegach funkcji przeżycia i ryzyka dla poszczególnych pacjentów są możliwe po wcześniejszym uzyskaniu $\hat{h}_0(t)$ i $\hat{S}_0(t)$, zgodnie z (11), (12) i (13). W tym celu całkowity czas obserwacji $[0, t_{j=10}] = [0, 50]$ podzielono na 10 podprzedziałów związanych z kolejnymi niepowodzeniami.

W każdym z nich obliczono $\xi(t_j)$, $\hat{h}_0(t_j)$ i $\hat{h}_0(t)$, co zestawiono w tabeli 5.

Tabela 5. Prawdopodobieństwo $\xi(t_j)$, $\hat{h}_0(t_j)$ i funkcje $\hat{h}_0(t)$ dla jednostki wzorcowej

Przedział czasowy	Prawdopodobieństwo przeżycia przez jednostkę wzorcową przedziału	Prawdopodobieństwo zajścia niepowodzenia dla jednostki wzorcowej w przedziale	Bazowa funkcja ryzyka
$[t_{j-1}, t_j)$	ξ_j	$\hat{h}_0(t_j)$	$\hat{h}_0(t_j)$
[0-2)	0,8203	0,1796	0,0898
[2-5)	0,8013	0,1987	0,0662
[5-10)	0,7558	0,2442	0,0488
[10-15)	0,7057	0,2943	0,0589
[15-20)	0,628	0,372	0,0744
[20-25)	0,3693	0,6307	0,1261
[25-28)	0,0675	0,9325	0,3108
[28-35)	0,0018	0,9981	0,1426
[35-40)	2,8E-06	0,9999	0,1999
[40-50)	0	1	0,1

Źródło: opracowanie własne.

Wyznaczone w tab. 10 estymatory $\hat{S}_0(t)$ i $\hat{h}_0(t)$ funkcji bazowych stanowią podstawę do określenia funkcji ryzyka $h_x(t)$ i przeżycia $S_x(t)$ dla pacjentów schakteryzowanych przez dowolne zmienne objaśniające zgodnie ze wzorami: $S_x(t_j) \equiv [\hat{S}_0(j)]^{\exp \beta x}$ i $h(x; t) = \hat{h}_0(t) \exp(\beta \cdot x)$. Tak więc, dla pacjenta leczonego

Tabela 6. Funkcje ryzyka i przeżycia dla pacjenta leczonego metodą A

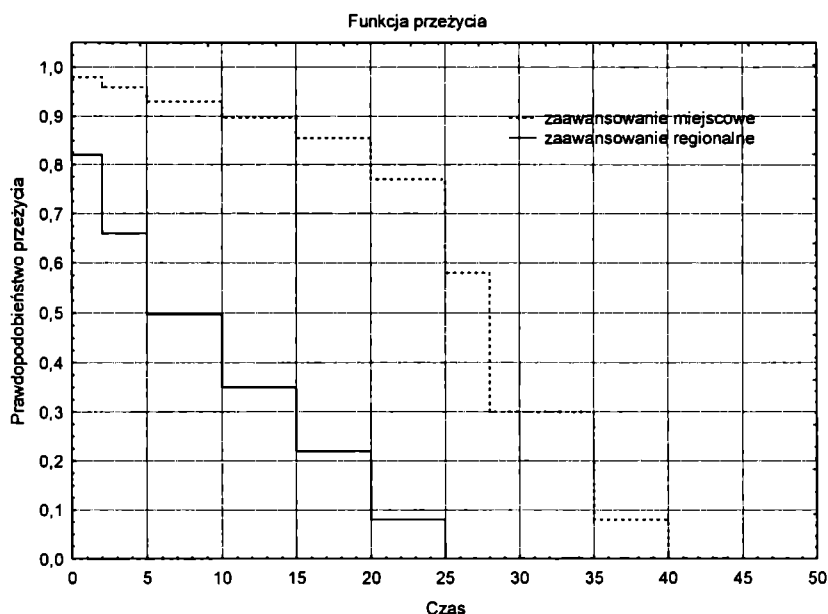
Przedział czasowy	Prawdopodobieństwo przeżycia przez jednostkę wzorcową przedziału	Pacjent leczony metodą A ($x_1 = 1$)			
		Choroba w zaawansowaniu miejscowym ($x_2 = 1$)		Choroba w zaawansowaniu regionalnym ($x_2 = 0$)	
		Funkcja przeżycia	Funkcja ryzyka	Funkcja przeżycia	Funkcja ryzyka
$[t_{j-1}, t_j)$	ξ_j	$S_x(t)$	$h(x; t)$	$S_x(t)$	$h(x; t)$
[0-2)	0,8204	0,9925	0,0034	0,9295	0,0331
[2-5)	0,8014	0,9842	0,0025	0,8566	0,0244
[5-10)	0,7558	0,9738	0,0018	0,7725	0,018
[10-15)	0,7057	0,961	0,0022	0,6793	0,0217
[15-20)	0,628	0,9442	0,0028	0,5722	0,0274
[20-25)	0,3693	0,9092	0,0048	0,3962	0,0465
[25-28)	0,0675	0,8209	0,0118	0,1465	0,1147
[28-35)	0,0018	0,6466	0,0054	0,0144	0,0526
[35-40)	2,8E-06	0,3986	0,0076	0,0001	0,0738
[40-50)	0	0	0,0038	0	0,0369

Źródło: opracowanie własne.

Tabela 7. Funkcje ryzyka i przeżycia dla pacjenta leczonego metodą B

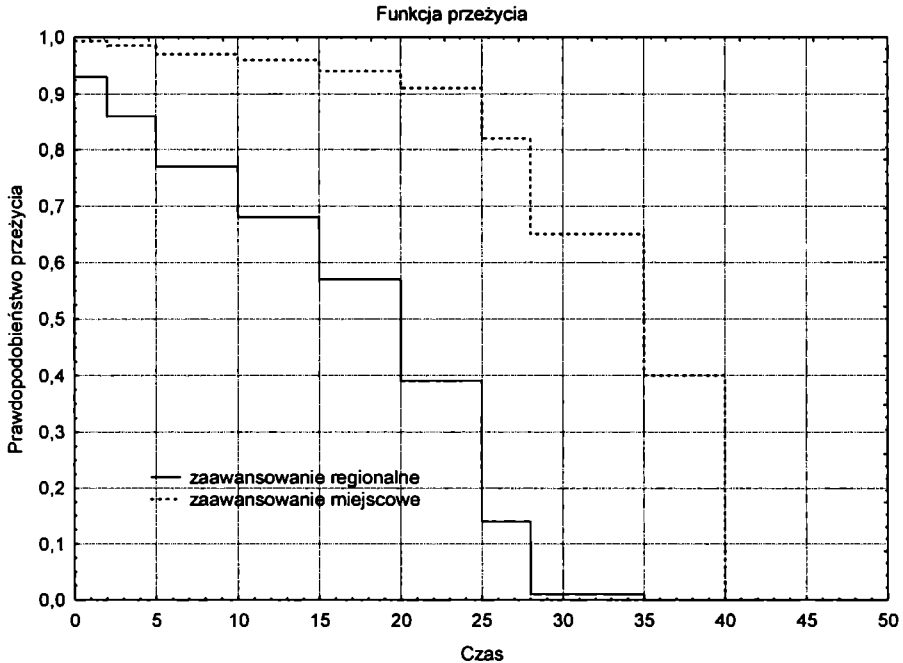
Przedział czasowy	Prawdopodobieństwo przeżycia przez jednostkę wzorcową przedziału	Pacjent leczony metodą B ($x_1 = 0$)			
		Choroba w zaawansowaniu miejscowym ($x_2 = 1$)		Choroba w zaawansowaniu regionalnym ($x_2 = 0$)	
		Funkcja przeżycia	Funkcja ryzyka	Funkcja przeżycia	Funkcja ryzyka
$[t_{j-1}, t_j]$	ξ_j	$S_x(t)$	$h(x;t)$	$S_x(t)$	$h(x;t)$
[0-2)	0,82036	0,9798	0,0092	0,82	0,0898
[2-5)	0,8013	0,9578	0,0068	0,6574	0,0662
[5-10)	0,7558	0,9306	0,005	0,4969	0,0488
[10-15)	0,7057	0,8979	0,006	0,3506	0,0588
[15-20)	0,628	0,856	0,0076	0,2202	0,0744
[20-25)	0,3693	0,7727	0,013	0,0813	0,1261
[25-28)	0,06746	0,5857	0,0319	0,0055	0,3108
[28-35)	0,0018	0,3067	0,0146	1,0E-05	0,1426
[35-40)	2,8E-06	0,0827	0,0205	2,9E-11	0,2
[40-50)	0	0	0,01	0	0,1

Źródło: opracowanie własne.



Rys. 2. Wyniki leczenia metodą B

Źródło: opracowanie własne.



Rys. 3. Wyniki leczenia metodą A

Źródło: opracowanie własne.

metodą A ($x_1 = 1$) będącego w zaawansowaniu miejscowym ($x_2 = 1$) powyższe funkcje są postaci: $h_{x_1=1, x_2=1} \equiv \hat{h}_0(t)e^{(-3,272)}$ i $S_{x_1=1, x_2=1}(t) \equiv [\hat{S}_0(t)]^{\exp(-3,272)}$, a w zaawansowaniu regionalnym ($x_2 = 0$): $h_{x_1=1, x_2=0} \equiv \hat{h}_0(t)e^{(-0,997)}$ i $S_{x_1=1, x_2=0}(t) \equiv [\hat{S}_0(t)]^{\exp(-0,997)}$. Pacjenta leczonego metodą B ($x_1 = 0$) będącego w zaawansowaniu miejscowym ($x_2 = 1$) charakteryzują funkcje: $h_{x_1=0, x_2=1} \equiv \hat{h}_0(t)e^{(-0,997)}$ i $S_{x_1=0, x_2=1}(t) \equiv [\hat{S}_0(t)]^{\exp(-0,997)}$, w zaawansowaniu regionalnym ($x_2 = 0$): $h_{x_1=0, x_2=0} \equiv \hat{h}_0(t)$ i $S_{x_1=0, x_2=0}(t) \equiv \hat{S}_0(t)$. Dokładne wartości funkcji przeżycia i ryzyka dla poszczególnych przypadków zestawiono w tab. 6 i 7 oraz zilustrowano na rys. 2 i 3.

Wnioski:

1. Obserwuje się bardzo szybki spadek wartości funkcji przeżycia dla pacjentów w zaawansowaniu regionalnym w stosunku do osób w zaawansowaniu miejscowym.
2. Metoda leczenia B jest skuteczniejsza od metody A.

7. Analiza wpływu wybranych czynników społeczno-ekonomicznych na wyniki leczenia chorych z rakiem głowy

Analizie poddano 122-osobową grupę pacjentów leczonych z powodu miejscowo zaawansowanego raka głowy z wykorzystaniem radioterapii. Oceniano czas przeżycia wolny od niepowodzeń leczenia T (ang. *Disease's Free Survival* – DFS) liczony od momentu zakończenia leczenia do wystąpienia wznowy miejscowej. W trakcie 5-letniej obserwacji odnotowano zajście 27 zdarzeń niepożądanych, czyli nawrotów miejscowych choroby. Obserwacje dla pozostałych pacjentów uznano za cenzurowane. Chorych różnicują zmienne objaśniające $X = \langle X_1, X_2, \dots, X_7 \rangle$ głównie o charakterze demograficznym oraz społeczno-ekonomicznym, których warianty przedstawiono w tab. 8.

Tabela 8. Zmienne objaśniające użyte do opisu pacjentów

Nazwa zmiennej	Oznaczenie zmiennej	Warianty zmiennej	Oznaczenie wariantów zmiennej
Płeć	X_1	kobieta	$x_1 = 0$
		mężczyzna	$x_1 = 1$
Wiek [lata]	X_2	zmienna ilościowa	x_2
Miejsce zamieszkania	X_3	wieś	$x_3 = 0$
		miasto	$x_3 = 1$
Dostęp do służby zdrowia	X_4	zły	$x_4 = 0$
		przeciętny	$x_4 = 1$
		dobry	$x_4 = 2$
Wykształcenie	X_5	wyższe	$x_5 = 0$
		pozostali	$x_5 = 1$
Aktywność zawodowa	X_6	pozostali	$x_6 = 0$
		pracujący	$x_6 = 1$
Stopień zaawansowania	X_7	T_1	$x_7 = 0$
		T_2	$x_7 = 1$

Źródło: opracowanie własne.

Korzystając ze wzoru (3), można wyznaczyć funkcję częściowej wiarygodności $L(\beta)$. Idea utworzenia tej funkcji jest identyczna jak w powyższych przykładach. Jest to iloczyn 27 prawdopodobieństw warunkowych związanych z niepowodzeniami leczenia poszczególnych osobników badanej grupy. Funkcja częściowej wiarygodności $L(\beta)$ przybiera postać wielomianu wysokiego stopnia, którego miejsca zerowe wyznaczono z użyciem pakietu Statistica 6.0 PL. Otrzymano następujące wartości współczynników regresji: $\beta_1 = 0,735$, $\beta_2 = 0,034$, $\beta_3 = 0,135$, $\beta_4 = -0,054$, $\beta_5 = 1,404$, $\beta_6 = -0,683$ oraz $\beta_7 = 0,987$. Następnie przeprowadzono redukcję zmiennych w modelu, stosując metodę krokową wsteczną.

Jako kryterium wykorzystano statystykę testową Walda i funkcje częściowej wiarygodności. Sposób konstrukcji optymalnego modelu jest szczegółowo opisany w literaturze [1; 5; 9]. Najlepszy model otrzymano dla następujących zmiennych objaśniających: X_7 , X_2 , X_5 . Funkcja $h(\mathbf{x}, t)$ dla opisywanego modelu ma postać:

$$h(\mathbf{x}, t) = h_0(t) \exp(1,06 X_7 + 1,40 X_5 + 0,04 X_2).$$

Pełne wyniki przedstawiono w tab. 9.

Tabela 9. Wyniki estymacji parametrów modelu

Zmienna	Ocena współczynnika β	Błąd średni szacunku współczynnika β	$\exp \beta$	Wartość statystyki Walda	Poziom istotności p dla statystyki Walda
X_5	1,40	0,748	4,05	3,489	0,062
X_2	0,04	0,02	1,04	4,102	0,043
X_7	1,06	0,5	2,88	4,486	0,034

Źródło: opracowanie własne w oparciu o dane kliniczne, z użyciem programu Statistica.

Oszacowane wartości parametrów β_2 , β_5 i β_7 pozwalają wyliczyć, zgodnie ze wzorem (14), współczynniki ryzyka HR dla występujących w modelu zmiennych objaśniających. Jeśli $x_{2j} = x_{2s}$ i $x_{5j} = x_{5s}$ oraz jeśli $x_{7j} = 1$ i $x_{7s} = 0$ to $HR = 2,88$. Natomiast, gdy $x_{2j} = x_{2s}$ i $x_{7j} = x_{7s}$ oraz gdy $x_{5j} = 1$ i $x_{5s} = 0$ to $HR = 4,05$. Dla jednostkowej różnicy w zmiennej wiek: $x_{2j} = x_{2s} + 1$, przy stałych wartościach pozostałych zmiennych wartość HR wynosi 1,04.

Analiza współczynników ryzyka wykazuje, że ryzyko niepowodzenia leczenia dla pacjenta w stopniu zaawansowania T_2 jest 2,88 razy większe niż dla pacjenta w zaawansowaniu T_1 , oczywiście przy takich samych wartościach pozostałych zmiennych. Osobę o wykształceniu wyższym charakteryzuje 4,05 razy większe ryzyko nawrotu choroby w stosunku do osoby z wykształceniem podstawowym, przy jednakowych wartościach pozostałych cech.

Dalsze analizy o konkretnych przebiegach funkcji przeżycia i ryzyka dla poszczególnych pacjentów są możliwe po wcześniejszym uzyskaniu $\hat{h}_0(t)$ i $\hat{S}_0(t)$, zgodnie z (11), (12) i (13). W tym celu całkowity czas obserwacji $[0, t_{j=27}] = [0, 60]$ podzielono na 27 podprzedziałów związanych z kolejnymi niepowodzeniami leczenia. W każdym z nich obliczono $\xi(t_j)$, $\hat{h}_0(t_j)$ i $\hat{h}_0(t)$. Wyznaczone estymatory $\hat{S}_0(t)$ i $\hat{h}_0(t)$ funkcji bazowych stanowią podstawę do określenia funkcji intensywności $h_x(t)$ i przeżycia $S_x(t)$ dla pacjentów scharakteryzowanych przez dowolne zmienne objaśniające zgodnie ze wzorami: $S_x(t_j) \equiv [\hat{S}_0(t_j)]^{\exp \beta^*}$ i $h(\mathbf{x}; t) = \hat{h}_0(t) \exp(\beta \cdot \mathbf{x})$. Oszacowano wartości funkcji przeżycia i ryzyka dla przykładowych pacjentów, co zestawiono w tab. 10. Ponadto dla przykładowych osobników badanej grupy wyznaczono wartość współczynnika ryzyka HR przyjmując

Tabela 10. Funkcje ryzyka i przeżycia dla przykładowych pacjentów

Przedział czasowy	Bazowa funkcja przeżycia	Bazowa funkcja ryzyka	Pacjent w wieku 60 lat ($x_2 = 60$); wykształcenie wyższe ($x_5 = 0$); zaawansowanie miejscowe T_1 ($x_7 = 0$);		Pacjent w wieku 60 lat ($x_2 = 60$); bez wykształcenia wyższego ($x_5 = 1$); zaawansowanie miejscowe T_1 ($x_7 = 0$);		Pacjent w wieku 60 lat ($x_2 = 60$); bez wykształcenia wyższego ($x_5 = 1$); zaawansowanie miejscowe T_2 ($x_7 = 1$);		Pacjent w wieku 80 lat ($x_2 = 80$); bez wykształcenia wyższego ($x_5 = 1$); zaawansowanie miejscowe T_2 ($x_7 = 1$);	
			Funkcja przeżycia	Funkcja ryzyka	Funkcja przeżycia	Funkcja ryzyka	Funkcja przeżycia	Funkcja ryzyka	Funkcja przeżycia	Funkcja ryzyka
$[t_{j-1}, t_j]$	$S_0(t)$	$\hat{h}_0(t)$	$S_x(t)$	$h(x; t)$	$S_x(t)$	$h(x; t)$	$S_x(t)$	$h(x; t)$	$S_x(t)$	$h(x; t)$
[0-1)	0,99	9,1E-05	0,99	0,001	0,99	0,004	0,99	0,01	0,97	0,29
[1-2)	0,99	9,3-05	0,99	0,001	0,99	0,004	0,98	0,01	0,95	0,29
[2-3)	0,99	9,4E-05	0,99	0,001	0,99	0,004	0,96	0,01	0,92	0,30
[3-4)	0,99	9,5E-05	0,99	0,001	0,98	0,004	0,95	0,01	0,90	0,30
[4-5)	0,99	9,9E-05	0,99	0,001	0,98	0,004	0,94	0,01	0,87	0,31
[5-6)	0,99	0,0001	0,99	0,001	0,97	0,004	0,93	0,01	0,85	0,32
[6-7)	0,99	0,0001	0,99	0,001	0,97	0,005	0,92	0,01	0,82	0,34
[7-8)	0,99	0,0001	0,99	0,001	0,96	0,005	0,90	0,01	0,80	0,35
[8-9)	0,99	0,0001	0,99	0,001	0,96	0,005	0,89	0,01	0,77	0,36
[9-10)	0,99	0,0001	0,99	0,001	0,95	0,005	0,87	0,01	0,74	0,39
[10-11)	0,99	0,0001	0,99	0,001	0,95	0,006	0,86	0,01	0,72	0,40
[11-12)	0,99	0,0001	0,99	0,001	0,94	0,006	0,85	0,02	0,69	0,41
[12-13)	0,99	0,0001	0,98	0,001	0,94	0,006	0,83	0,02	0,66	0,42
[13-14)	0,99	0,0001	0,98	0,001	0,93	0,006	0,82	0,02	0,64	0,44
[14-15)	0,99	0,0001	0,98	0,001	0,93	0,006	0,80	0,02	0,61	0,45
[15-16)	0,99	0,0001	0,98	0,0015	0,92	0,006	0,79	0,02	0,59	0,45
[16-17)	0,99	0,0001	0,98	0,0015	0,91	0,006	0,77	0,02	0,56	0,46
[17-18)	0,99	0,0001	0,98	0,0016	0,91	0,006	0,76	0,02	0,54	0,46
[18-19)	0,99	0,0001	0,97	0,0016	0,9	0,007	0,74	0,02	0,52	0,47
[19-20)	0,99	0,0001	0,97	0,0016	0,9	0,007	0,73	0,02	0,50	0,48
[20-22)	0,99	7,9E-05	0,97	0,0008	0,89	0,003	0,71	0,01	0,47	0,25
[22-23)	0,99	0,0002	0,97	0,0018	0,88	0,007	0,70	0,02	0,45	0,51
[23-25)	0,99	8,205	0,97	0,0009	0,88	0,004	0,69	0,01	0,43	0,26
[25-30)	0,99	3,5E-05	0,97	0,0004	0,87	0,002	0,67	0,004	0,41	0,11
[30-50)	0,99	1,2-05	0,96	0,0001	0,86	0,001	0,65	0,001	0,38	0,04
[50-51)	0,99	0,0002	0,96	0,0027	0,85	0,011	0,63	0,03	0,36	0,78
[51-54)	0,99	9,6E-05	0,96	0,0010	0,84	0,004	0,61	0,01	0,33	0,30

Źródło: opracowanie własne w oparciu o dane kliniczne, przy użyciu programów Statistica i Excel.

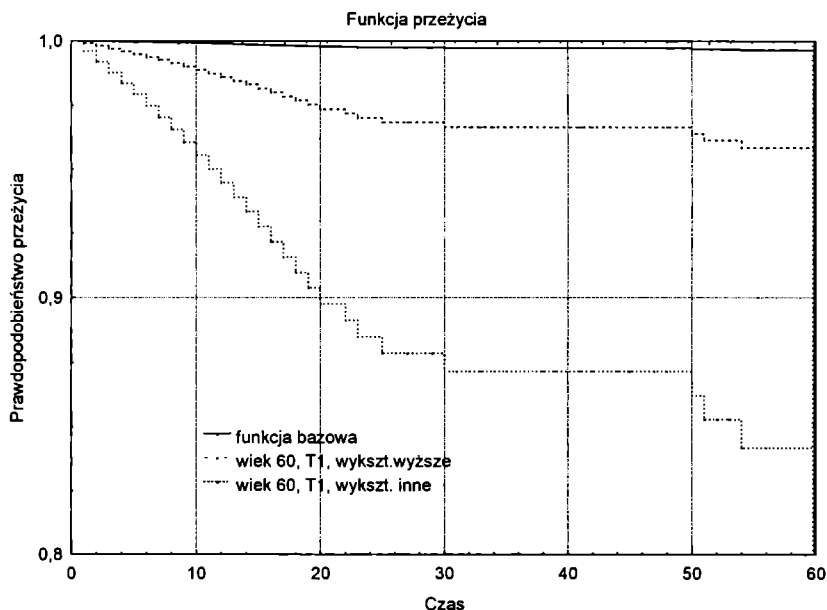
jako punkt odniesienia wartość bazowej funkcji ryzyka, tzn. $x_{2s} = x_{5s} = x_{7s} = 0$. Wyniki zawarto w tab. 11.

Tabela 11. Współczynniki ryzyka dla przykładowych pacjentów

Pacjent	Wartość x_2	Wartość x_5	Wartość x_7	Wartość HR
j -ty	60	0	0	11,023
s -ty	0	0	0	
j -ty	60	1	0	44,701
s -ty	0	0	0	
j -ty	60	1	1	129,024
s -ty	0	0	0	
j -ty	80	1	1	287,149
s -ty	0	0	0	

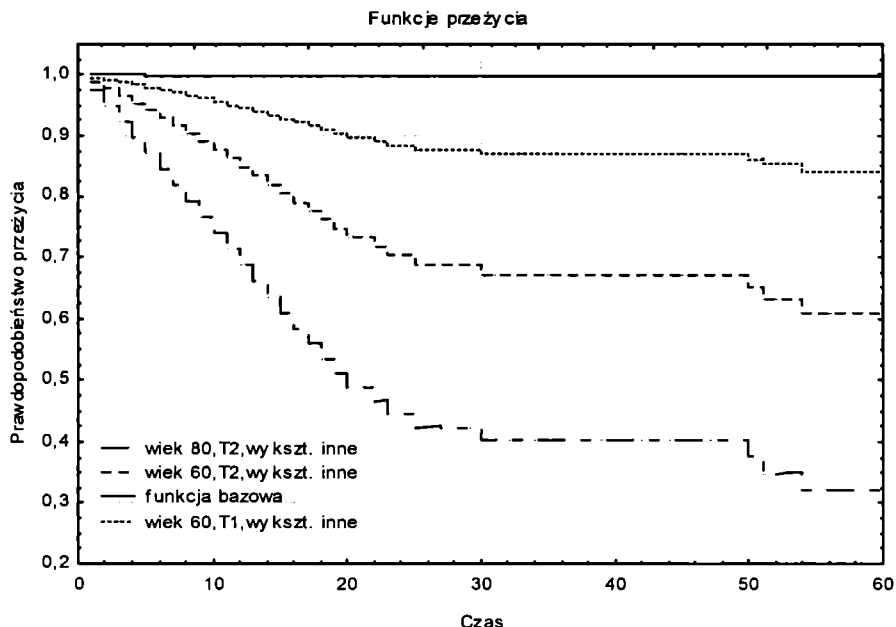
Źródło: opracowanie własne na podstawie danych klinicznych.

Graficzną ilustrację przebiegu funkcji przeżycia dla przykładowych pacjentów przedstawiono na rys. 4 i 5.



Rys. 4. Krzywe przeżycia dla pacjentów z wykształceniem wyższym i innym

Źródło: opracowanie własne.



Rys. 5. Krzywe przeżycia dla pacjentów w różnym wieku i różnych stopniach zaawansowania choroby nowotworowej

Źródło: opracowanie własne.

Wnioski:

1. W badanej grupie pacjentów łączny wpływ na wyniki leczenia mają następujące zmienne: wiek, stopień zaawansowania choroby i wykształcenie.
2. Efekt oddziaływania poszczególnych zmiennych jest wyrażony przez współczynniki ryzyka HR osiągające największą wartość dla wykształcenia.
3. Zaobserwowano bardzo szybki spadek przebiegu krzywych przeżycia wraz ze wzrostem wieku i stopnia zaawansowania nowotworu.

8. Analiza wpływu wybranych czynników na wyniki leczenia chorych z rakiem szyjki macicy

Analizie poddano 107-osobową grupę pacjentek leczonych operacyjnie z powodu raka szyjki macicy. Oceniano czas przeżycia T liczony od momentu operacji do dnia śmierci. Uwzględniono tylko zgony spowodowane przez leczoną chorobę nowotworową. W trakcie 5-letniej obserwacji odnotowano zajście 23 takich zdarzeń niepożądanych. Przypadki pozostałe uznano jako cenzurowane. Chorych różnicują zmienne objaśniające $X = \langle X_1, X_2, \dots, X_8 \rangle$ głównie o charakterze medycznym, ale także demograficzne oraz społeczne. Ich warianty przedstawiono w tab. 12.

Tabela 12. Zmienne objaśniające użyte do opisu pacjentek

Nazwa zmiennej	Oznaczenie zmiennej	Warianty zmiennej	Oznaczenie wariantów zmiennej
Wiek [lata]	X_1	zmienna ilościowa	x_1
Miejsce zamieszkania	X_2	wieś	$x_2 = 0$
		miasto	$x_2 = 1$
Wykształcenie	X_3	wyższe	$x_3 = 0$
		pozostali	$x_3 = 1$
Metoda leczenia	X_4	standardowa	$x_4 = 0$
		inna	$x_4 = 1$
Przerzuty do węzłów chłonnych	X_5	brak	$x_5 = 0$
		obecne	$x_5 = 1$
Histopatologia	X_6	rak płaskonabłonkowy	$x_6 = 0$
		inne	$x_6 = 1$
Stopień zróżnicowania nowotworu	X_7	dobrze zróżnicowany	$x_7 = 0$
		niezróżnicowany	$x_7 = 1$
Stopień zaawansowania miejscowego raka	X_8	I A	$x_8 = 0$
		I B	$x_8 = 1$
		II	$x_8 = 2$

Źródło: opracowanie własne.

Postępując analogicznie jak w poprzednim przykładzie, skonstruowano optymalny model z następującymi zmiennymi objaśniającymi: X_5 i X_8 . Funkcja $h(x, t)$ dla opisywanego modelu ma postać:

$$h(x, t) = h_0(t) \exp(0,98 X_5 + 1,02 X_8).$$

Pełne wyniki przedstawiono w tab. 13.

Tabela 13. Wyniki estymacji parametrów modelu

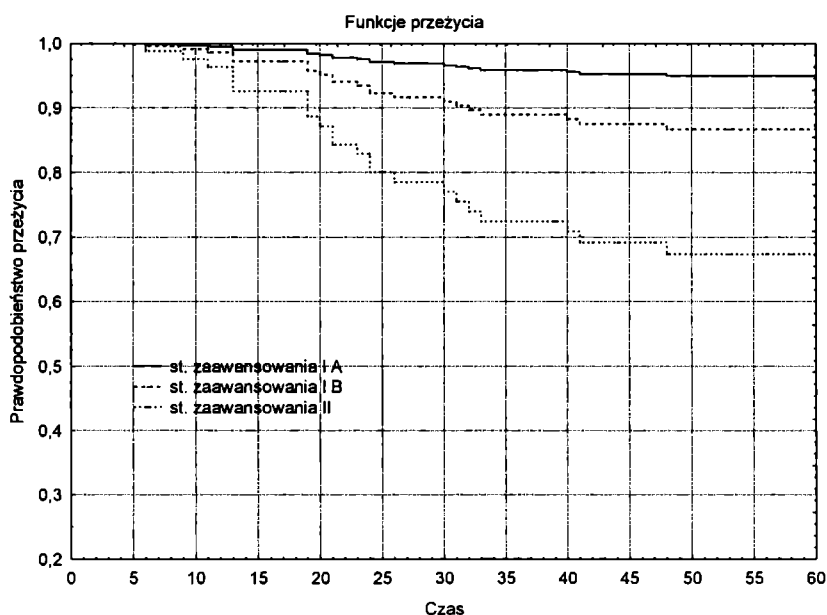
Zmienna	Ocena współczynnika β	Błąd średni szacunku współczynnika β	$\exp \beta$	Wartość statystyki Walda	Poziom istotności p dla statystyki Walda
X_5	0,98	0,439	2,66	4,99	0,025
X_8	1,02	0,386	2,76	6,928	0,008

Źródło: opracowanie własne w oparciu o dane kliniczne, z użyciem programu Statistica.

Oszacowane wartości parametrów β_5 i β_8 pozwalają wyliczyć, zgodnie ze wzorem (14), współczynniki ryzyka HR dla odpowiednich zmiennych objaśniających. Jeśli $x_{8j} = x_{8s}$ oraz jeśli $x_{5j} = 1$ i $x_{5s} = 0$, to $HR = 2,66$. Natomiast gdy $x_{5j} = x_{5s}$ oraz gdy $x_{8j} = 2$ i $x_{8s} = 1$, to $HR = 2,76$. W przypadku cechy x_5 oznacza to, że ryzyko zajścia zdarzenia niepożądanego jest 2,66 razy większe dla chorych z obecnością przerzutów do węzłów chłonnych w stosunku do chorych z wolnymi węzłami. Drugą zmienną warunkującą wyniki leczenia jest stopień zaawansowania nowotworu. Ryzyko śmierci chorej z rakiem w stopniu IB jest 2,76 razy wyższe niż pacjentki w stopniu IA, podobnie jak chorej w stopniu II w stosunku do pacjentki

w stopniu IB. Najgorzej rokuje chora w stopniu II i z obecnością przerzutów do węzłów chłonnych, ma 20,5 raza większe ryzyko śmierci w stosunku do chorej w stopniu IA bez przerzutów.

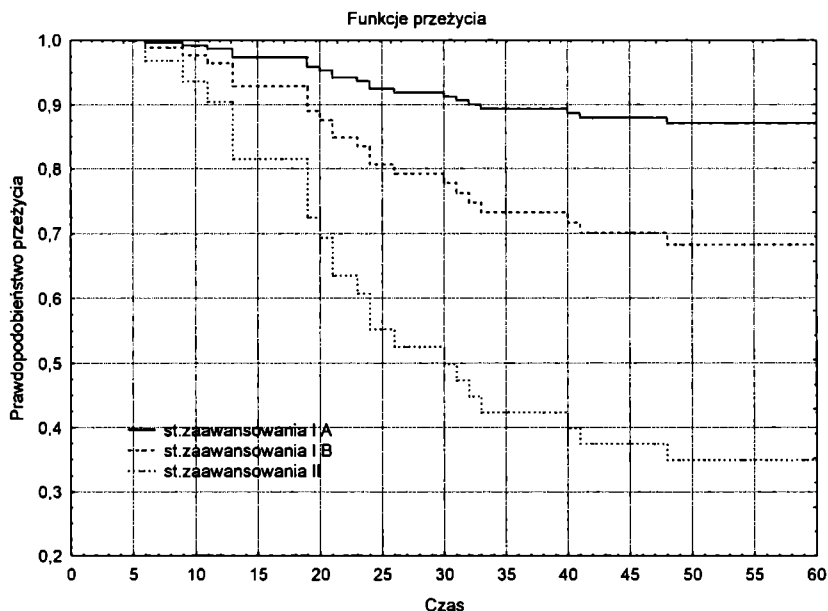
Badanie konkretnych przebiegów funkcji przeżycia i ryzyka dla poszczególnych pacjentów jest możliwe po wcześniejszym oszacowaniu $\hat{h}_0(t)$ i $\hat{S}_0(t)$, zgodnie z (11), (12) i (13). Wyznaczone estymatory stanowią podstawę do określenia funkcji intensywności $h_x(t)$ i przeżycia $S_x(t)$ dla pacjentów scharakteryzowanych przez dowolne zmienne objaśniające zgodnie ze wzorami: $S_x(t_j) \equiv [\hat{S}_0(j)]^{\exp \beta x}$ i $h(x;t) = \hat{h}_0(t) \exp(\beta \cdot x)$. Oszacowano wartości funkcji przeżycia dla przykładowych pacjentek celem ilustracji wpływu zmiennych objaśniających, co przedstawiono na rys. 6 i 7.



Rys. 6. Funkcje przeżycia dla pacjentek bez przerzutów do węzłów chłonnych

Źródło: opracowanie własne.

Na rys. 6 przedstawiono przebieg funkcji przeżycia dla chorych bez przerzutów do węzłów chłonnych. Wykres ilustruje negatywny wpływ wyższego stopnia zaawansowania na prawdopodobieństwo przeżycia. Krzywa dla chorej w stopniu IA jest równoznaczna z bazową funkcją przeżycia. Poniższy wykres zawiera funkcje przeżycia dla pacjentek w różnych stadiach choroby i z obecnością przerzutów w węzłach. Najmniejszą szansę na wyleczenie w dłuższym przedziale czasowym mają chore w stopniu choroby II.



Rys. 7. Funkcje przeżycia dla pacjentek z przerzutami do węzłów chłonnych

Źródło: opracowanie własne.

Wnioski:

1. Wpływ na wyniki leczenia w badanej grupie mają następujące zmienne: stopień zaawansowania miejscowego choroby i obecność przerzutów do węzłów chłonnych.
2. Gorzej rokują pacjentki w wyższych stopniach zaawansowania miejscowego i z przerzutami, co ilustrują wartości współczynników ryzyka HR .
3. Najgorzej rokująca grupa to chore z II stopniem zaawansowania i z obecnością przerzutów do węzłów chłonnych.

9. Podsumowanie – wnioski końcowe

1. Model Coxa umożliwia wskazanie czynników wpływających na ryzyko wystąpienia zdarzenia niepożądanego przez obliczenie wartości współczynnika ryzyka HR .

2. Pełną informację o przebiegu funkcji przeżycia i ryzyka dla osobników scharakteryzowanych wektorem cech $x = \langle x_1, x_2, \dots, x_m \rangle$ można uzyskać po wcześniejszym oszacowaniu (np. za pomocą estymatorów Kalbfleischa–Prentice’a) funkcji bazowych $h_0(t)$ i $S_0(t)$.

3. Czynniki demograficzne, społeczne i ekonomiczne mogą mieć oprócz cech o charakterze medycznym wpływ na wyniki leczenia niektórych chorób nowo-

tworowych. Wykazano wpływ wieku i poziomu wykształcenia na rezultaty terapii chorych z rakiem głośni.

Literatura

- [1] Collet D., *Modelling Survival Data in Medical Research*, Chapman & Hall, Boca Raton 1994.
- [2] Cox D.R., *Regression Models and Life-Tables*, „Journal of the Royal Statistical Society” 1972 34, 187-220.
- [3] Frątczak E., *Analiza historii zdarzeń – elementy teorii, wybrane przykłady zastosowań z wykorzystaniem pakietu TDA*, SGH, Warszawa 1997.
- [4] Kalbfleisch J.D., Prentice R.L., *Marginal Likelihoods Based On Cox's Regression and Life Model*, „Biometrika” 1973, 60, 267-78.
- [5] Kalbfleisch J.D., Prentice R.L., *The Statistical Analysis of Failure Time Data*, Wiley-Interscience, 2002.
- [6] Kaplan E.L., Meier P., *Nonparametric Estimation from Incomplete Observations*, „Journal of the American Statistical Association” 1953, 53, 457-81.
- [7] Kleinbaum D.G., *Survival Analysis*, Springer, New York 1997.
- [8] Miller R.G., *Survival Analysis*, John Wiley & Sons, New York 1988.
- [9] Ostasiewicz S. (red.), *Metody oceny i porządkowania ryzyka w ubezpieczeniach życiowych*, AE, Wrocław 2000.
- [10] Stanisław A., *Przystępny kurs statystyki*, t. 2, StatSoft Polska, 2000.
- [11] *Statistica Manual*, t. 3, StatSoft Polska, 1997.

ESTIMATION OF COX MODEL BASED ON THE RESULTS OF MALIGNANT TUMOURS THERAPY

Summary

The paper discusses the proportional hazard model by Cox describing the phenomena in which the unexpected events (e.g. deaths) exist in the disjoint time intervals. Kalbfleische–Prentice estimators have been to conformed such a model variation, which determines the baseline hazard function $h_0(t)$ and the baseline survivals function $S_0(t)$. In practice it enables to determine the survival $S_x(t)$ and hazard $h(x; t)$ functions for the researched units varied with the vector of endo- and exogenous features. The discussed functions have been estimated on the medical examples that include patients suffering from malignant tumours. The influence of demographical, social, and economic features on the therapy results has been discussed.