

*dr hab. Mieczysław Jan Król*¹

Katedra Metod Ilościowych i Informatyki Gospodarczej, Wydział Ekonomii
Uniwersytet Rzeszowski

*lek. med. Piotr Sewer Król*²

Kliniczny Zakład Radioterapii
Wojewódzki Szpital Specjalistyczny w Rzeszowie

Model proporcjonalnego hazardu Coxa jako statystyczna metoda wyznaczania determinant jakości życia w grupie chorych leczonych onkologicznie

WSTĘP

Koncepcje dotyczące podstawowych zasad zarządzania trwałym rozwojem w społeczeństwie informacyjnym rozwarstwiają się na dwa nurty. W pierwszym są skupione na racjonalnym wykorzystaniu zasobów naturalnych do zrealizowania celów gospodarczych, a w drugim odnoszą się do ekosystemu, zdrowia i uwagunkowań dotyczących godnego życia. Osnową koncepcji w obu nurtach są zagadnienia dotyczące zaspakajania potrzeb determinujących odpowiedni poziom jakości życia, co wiąże się nie tylko ze wzrostem dochodów osobistych, ale również z edukacją i wychowaniem, dostępem do wiedzy i informacji, prawem do pracy, możliwością korzystania z dóbr kultury, rekreacji i turystyki, warunkami mieszkaniowymi, gwarancją bezpieczeństwa w tym bezpieczeństwa socjalnego oraz szeroko pojmowaną ochroną zdrowia [Hibner, 2003].

Jakość życia jest pojęciem relatywnym, więc jej badanie w systemach ekonomiczno-społecznych uzależnianiem dobrobytu od dochodów może generować fałszywe wnioski [Skrzypek (red.), 2007, s. 59–79]. Zatem do badań oraz analiz porównujących jakość życia w różnych grupach społecznych tworzone są wielorakie indeksy agregatowe. Do znanych i wykorzystywanych należą indeksy charakter-

¹ Adres korespondencyjny: Katedra Metod Ilościowych i Informatyki Gospodarczej, Wydział Ekonomii, Uniwersytet Rzeszowski, ul. Ćwiklińskiej 2, 35-601 Rzeszów, e-mail: krolmeav@city.net.pl, tel. 17 872 16 99.

² Adres korespondencyjny: Kliniczny Zakład Radioterapii, Wojewódzki Szpital Specjalistyczny w Rzeszowie, ul. Szopena 2, 35-055 Rzeszów, e-mail: krolmeav@city.net.pl, tel. 17 86 66 400.

ryzujące ilościowo warunki życia (LCI – *Living Conditions Index*), a także jakość życia (QOL – *Quality of Life*) oraz rozwój człowieka (HDI – *Human Development Index*) [Biernacki, 2009, s. 55–96]. W konstrukcji indeksów wykorzystywane są mierniki odnoszące się do triady zdrowie – wykształcenie – zamożność, w której występuje bogaty wachlarz przeróżnych aspektów życia mocno ze sobą skorelowanych. Logiczny sens triady polega na tym, że zdrowy obywatel wydajniej pracuje i racjonalniej wypoczywa, co pozytywnie wpływa na jego dochody. Ponadto, zdrowy i zamożny obywatel ma większe możliwości zdobywania wiedzy i podnoszenia swoich kwalifikacji. Natomiast dobre wykształcenie zwiększa szanse znalezienia dobrze płatnej pracy oraz zwiększa świadomość zdrowotną wyrażającą się w staraniach zachowania dobrego stanu zdrowia poprzez racjonalne odżywianie, czy też efektywne wykorzystanie czasu wolnego. Zamożny obywatel w szerokim zakresie może korzystać z oferty rynku edukacyjnego, rynku zdrowia, co podnosi na wyższy poziom jego jakość życia.

Krytycy wskaźników agregatowych dopatrują się w nich wad dotyczących arbitralnego doboru elementów składowych, a także subiektywnego określania ważności i dowolnego ustalania minimalnych oraz maksymalnych progów dla poziomów zaspakajania przeróżnych potrzeb człowieka. Do wnikliwych badań, zwłaszcza zjawisk przebiegających dynamicznie w konkretnych obszarach i na wybranych populacjach, stosuje się wskaźniki szczegółowe dotyczące np. czasu życia lub momentu wystąpienia zdarzenia niepożądanego u jednostki funkcjonującej w określonej kohorcie (tj. grupie) społecznej [Collet, 1994]. Wskaźniki związane z czasem życia i występujące w modelu proporcjonalnego ryzyka (hazardu) Coxa [Cox, 1972, s. 187–220] znajdują merytoryczne uzasadnienie w zaawansowanych teoriach matematycznych i są wykorzystywane nie tylko do badania zjawisk związanych z jakością życia [Ostasiewicz (red.), 2000], ale również do badań niezawodnego funkcjonowania skomplikowanych urządzeń technicznych [Kopociński, 1973]. W artykule model Coxa jest wykorzystany do wyodrębnienia cech ekonomiczno-społecznych determinujących czas życia pacjentów leczonych na chorobę nowotworową. Wartości liczbowe dla parametrów modelu zostały wyznaczone metodologią znaną w literaturze [Miller, 1988] z wykorzystaniem numerycznego oprogramowania *STATISTICA PL* [Instrukcja..., 1997] w oparciu o dane uzyskane dla badanej grupy w Klinicznym Zakładzie Radioterapii Wojewódzkiego Szpitala Specjalistycznego w Rzeszowie.

IDEA MATEMATYCZNEGO UJĘCIA MODELU PROPORCJONALNEJ INTENSYWNOŚCI COXA

Liczebność $n=n(t)$ obserwowanej oraz leczonej grupy (kohorty) Ω pacjentów cierpiących na pewną chorobę jest nierosnącą, prawostronnie ciągłą i przedziałami stałą funkcją czasu $t, t \geq 0$. Na wartości tej funkcji wpływają różne zdarzenia,

w tym niepożądane (np. nawrót choroby lub zgon pacjenta) oraz zdarzenia cenzurowane³. Jeśli $n_0=n(t_0=0)$ oznacza początkową liczebność kohorty, a $n_j=n(t_j)$ jej liczebność w momencie t_j , w którym zaistniało zdarzenie niepożądane⁴, przy czym $0=t_0<t_1<t_2<\dots<t_j<\dots<t_r<\dots$, to nierówność $n_{j-1}-n_j>1$ oznacza wystąpienie w przedziale (t_{j-1}, t_j) co najmniej jednego zdarzenia ocenzonego, a $\lambda(j)=n_j+1$ liczebność kohorty tuż przed zaistnieniem zdarzenia niepożądanego w chwili t_j ($j=1, 2, \dots, r$).

Pacjentów różnicuje zbiór cech $\{X_1, X_2, \dots\}$ z którego, kierując się kryteriami opartymi na związkach przyczynowo-skutkowych, wybierane jest tylko m spośród nich do objaśniania czasu życia T , $T \geq 0$, każdej jednostki z kohorty.

Cechy te przedstawia wektor $X = \langle X_1, X_2, \dots, X_m \rangle$ ⁵. W modelu przy założeniu znajomości rozkładu prawdopodobieństwa zmiennej losowej T dla pacjenta $\omega \in \Omega$ scharakteryzowanego konkretnymi wartościami $x(\omega) = \langle x_1(\omega), x_2(\omega), \dots, x_m(\omega) \rangle$ cech, jest definiowana funkcja ryzyka (hazardu) $h(x(\omega), t)$ jako intensywność prawdopodobieństwa wystąpienia u niego zdarzenia niepożądanego w chwili t , a mianowicie⁶:

$$h(x(\omega), t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t} = \frac{f(x(\omega), t)}{S(x(\omega), t)}, \quad (1)$$

gdzie: $F(x(\omega), t)$ jest dystrybucją rozkładu prawdopodobieństwa czasu T życia tegoż pacjenta, $S(x(\omega), t) = 1 - F(x(\omega), t)$ jego funkcją przeżycia okresu czasu $[0, t)$ a $f(x(\omega), t) = dF(x(\omega), t)/dt$ funkcją gęstości prawdopodobieństwa. W pracy [Cox, 1972, s. 187–220] Cox zaproponował następującą postać analityczną dla funkcji ryzyka:

$$h(x(\omega), t) = h(0, t) \cdot \exp[\beta \bullet x(\omega)], \quad (2)$$

gdzie: $\beta \bullet x(\omega) = \beta_1 x_1(\omega) + \beta_2 x_2(\omega) + \dots + \beta_m x_m(\omega)$ jest iloczynem skalarnym wektora $x(\omega)$ cech charakteryzujących pacjenta $\omega \in \Omega$ oraz wektora parametrów $\beta = \langle \beta_1, \beta_2, \dots, \beta_m \rangle$ adekwatnych dla całej kohorty. Natomiast czynnik $h(0, t)$ jest funkcją ryzyka bazowego zależną tylko od czasu, a drugi, niezależny od czasu t ($t \geq 0$) czynnik $\exp[\beta \bullet x(\omega)]$ jest związany z pacjentem $\omega \in \Omega$ poprzez wektor $x(\omega) = \langle x_1(\omega), x_2(\omega), \dots, x_m(\omega) \rangle$ wartości cech. Z równań (1) oraz (2) wynika analityczna postać dla prawdopodobieństwa wystąpienia zdarzenia niepożądanego u pacjenta $\omega \in \Omega$ w przedziale czasowym $[t, t + \Delta t)$ przy założeniu, że nie wystąpiło ono w okresie $[0, t)$, a mianowicie:

³ Cenzurowanie, to utrata jednostki z kohorty bez odnotowania faktu, że na tej jednostce zaistniało zdarzenie niepożądane.

⁴ W momencie t_j pojawia się tylko jedno zdarzenie niepożądane (model prosty).

⁵ W podstawowym modelu Coxa zakłada się niezależność zmiennych objaśniających X_1, X_2, \dots, X_m od czasu t , ale we współczesnej literaturze rozpatrywane są modele bez tego ograniczenia zob. np.: [Kalbfleish, Prentice, 2002].

⁶ t_j na końcu przedziału czasowego $[0, t]$.

$$P(t \leq T < t + \Delta t \mid T \geq t) = h(0, t) \cdot \exp[\beta \bullet x(\omega)] \cdot \Delta t + o(\Delta t), \quad (3)$$

gdzie: $[o(\Delta t)/\Delta t] \rightarrow 0$ o ile $\Delta t \rightarrow 0$. Zatem w modelu Coxa jednostki posiadające te same cechy mają tą samą funkcję ryzyka (hazardu), a relatywne odniesienie funkcji ryzyka pacjenta ω_j względem funkcji pacjenta ω_s nie zależy od czasu. Jeśli wprowadzimy oznaczenia $x(j) = x(\omega_j) = \langle x_1(j), x_2(j), \dots, x_m(j) \rangle$, $x(s) = x(\omega_s) = \langle x_1(s), x_2(s), \dots, x_m(s) \rangle$ cech charakteryzujących pacjentów $\omega_j, \omega_s \in \Omega$ skracając zapis, to warunki te mają następującą postać matematyczną:

$$x(j) = x(s) \Rightarrow h(x(j), t) = h(x(s), t) \quad (4)$$

oraz

$$HR_{j,s} = \frac{h(x(j), t)}{h(x(s), t)} = \exp[\beta \bullet (x(j) - x(s))] = \exp \sum_{i=1}^m \beta_i (x_i(j) - x_i(s)). \quad (5)$$

Ponieważ $HR_{j,s}$ jest współczynnikiem proporcjonalności dla funkcji ryzyka (hazardu) pacjentów $\omega_j, \omega_s \in \Omega$, więc naturalnie staje się on wskaźnikiem do porównań ryzyka (ang. *hazard ratio*) [Kleinbaum, 1997] wystąpienia zdarzenia niepożądanego na dwóch różnych jednostkach kohorty. Jeśli $HR_{j,s} > 1$, to z uwagi na równość (5) jest $h(x(j), t) = HR_{j,s} \cdot h(x(s), t) > h(x(s), t)$ co oznacza, że prawdopodobieństwo zaistnienia zdarzenia niepożądanego u pacjenta ω_j jest większe od prawdopodobieństwa zaistnienia tego samego zdarzenia, ale u pacjenta ω_s . Fakt, że iloraz funkcji ryzyka dla dwóch jednostek o różnych wartościach cech objaśniających nie zależy od czasu jest wystarczającym warunkiem do stwierdzenia, że semiparametryczny model regresji z nieujemną funkcją hazardu (ryzyka) Coxa jest modelem proporcjonalnej intensywności dla prawdopodobieństwa wystąpienia zdarzenia niepożądanego w chwili t , $t > 0$.

Prawdopodobieństwo wystąpienia zdarzenia niepożądanego u pacjenta $\omega \in \Omega$ w dowolnym momencie przedziału czasowego $[0, t)$ jest określone całką.

$$H(x(\omega), t) = \int_0^t h(x(\omega), \tau) d\tau$$

i jest nazywane skumulowaną funkcją ryzyka (lub hazardu). Wykorzystując w całej równość (1) skumulowana funkcja ryzyka po przekształceniach przyjmuje postać:

$$H(x(\omega), t) = H(0, t) \cdot \exp(\beta \bullet x) = -\ln S(x(\omega), t), \quad (6)$$

gdzie: $H(0, t) = \int_0^t h(0, \tau) d\tau$ jest tzw. skumulowanym ryzykiem bazowym. Stąd dla funkcji przeżycia $S(x(\omega), t) = 1 - F(x(\omega), t)$ przedziału czasowego $[0, t)$ przez jednostkę $\omega \in \Omega$ mającą wektor $x(\omega) = \langle x_1(\omega), x_2(\omega), \dots, x_m(\omega) \rangle$ wartości cech, wynika wzór:

$$S(x(\omega), t) = [S(0, t)]^{\exp[\beta \bullet x(\omega)]}, \quad (7)$$

w którym $S(0, t) = \exp[-H(0, t)]$ jest bazową funkcją przeżycia przedziału czasowego $[0, t)$.

ODNIESIENIA DO ESTYMACJI PARAMETRÓW MODELU

Ze względu na występowanie cenzurowania danych⁷, proces estymowania wartości liczbowych istotnych dla modelu parametrów $\beta = \langle \beta_1, \beta_2, \dots, \beta_m \rangle$ jest oparty na prawdopodobieństwach warunkowych $P(A(t_j) | B(t_j))$, w których zdarzenia losowe dotyczą: $A(t_j)$ – zaistnienia zdarzenia niepożądanego na jednostce $\omega_j \in \Omega$ w momencie t_j oraz $B(t_j)$ – wystąpienia tylko jednego zdarzenia niepożądanego w chwili t_j na jednostce $\omega \in \Lambda(j) \subseteq \Omega$, gdzie $\Lambda(j)$ jest podzbiorem złożonym z tych jednostek kohorty Ω , które żyją i są obserwowane tuż przed momentem t_j . Momenty zaistnienia kolejnych zdarzeń niepożądanych $0 < t_1 < t_2 < \dots < t_j < \dots < t_r$ ($j=1, 2, \dots, r$) tworzą niemalejący ciąg $\{t_j; j=1, 2, \dots, r\}$. Związek generowania podzbiorów $\Lambda(j) \subseteq \Omega$ przez zdarzenia niepożądane i przez obserwacje ocenzone w przedziale czasowym $[0, t_j]$ jest oczywisty. Dla modelu proporcjonalnego hazardu Coxa funkcja warunkowej wiarygodności ma postać [Miller, 1988]:

$$L(\beta) = \prod_{j=1}^r \left(\frac{\exp(\beta \cdot x(j))}{\sum_{s \in \Lambda(j)} \exp(\beta \cdot x(s))} \right)^{\delta_j}, \quad (8)$$

w której δ_j jest wskaźnikiem cenzurowania⁸. Liczbowe wartości parametrów $\beta_1, \beta_2, \dots, \beta_m$ charakteryzujących kohortę jako całość otrzymuje się rozwiązując układ równań powstały z przyrównania gradientu funkcji $L(\beta)$ do zera, tj.:

$$\text{grad}L(\beta) = \langle \partial L(\beta) / \partial \beta_1, \partial L(\beta) / \partial \beta_2, \dots, \partial L(\beta) / \partial \beta_m \rangle = 0$$

zgodnie z zasadami matematycznymi dla zadań ekstremalnych $L(\beta) \rightarrow \text{MAX}$.

Model optymalny otrzymuje się stosując znaną w literaturze metodę krokową wsteczną z kryterium będącym statystyką Walda. Kolejno eliminuje się te cechy, dla których komputerowy poziom istotności jest najwyższy [Ostasiewicz, 2000]. Dla tak powstałych modeli statystyką chi-kwadrat sprawdza się ich dopasowanie do danych [Stanisz, 2000]. Za optymalny uznawany jest ten, dla którego wskaźnik dopasowania jest najlepszy. Estymatory $S^{\wedge}(t)$ i $h^{\wedge}(t)$ dla funkcji bazowych $S(0, t)$ i $h(0, t)$ stanowią podstawę do określenia funkcji intensywności ryzyka $h(x(\omega), t)$ i funkcji przeżycia $S(x(\omega), t)$ dla pacjentów scharakteryzowanych przez wartości determinant, czyli istotnych zmiennych objaśniających $x(\omega) = \langle x_1(\omega), x_2(\omega), \dots, x_m(\omega) \rangle$ zgodnie ze wzorami:

$$S(x(\omega), t) \approx [S^{\wedge}(0, t)] \exp[\beta \cdot x(\omega)], \quad h(x(\omega), t) \approx h^{\wedge}(t) \cdot \exp[\beta \cdot x(\omega)]. \quad (9)$$

Skonstruowano wiele estymatorów h^{\wedge} oraz S^{\wedge} dla funkcji bazowych (Breslow, Tsatis, Link, Kaplan i Meier) [Miller, 1998], ale przy odpowiednich ograniczeniach np. estymatory Kaplana-Meiera [Kaplan, Meier, 1958, s. 457–81] wymagają spełnienia założenia, że z tym samym prawdopodobieństwem pojawia się zdarzenie niepożądane na każdej jednostce z kohorty. Spełnienie tego warunku

⁷ W badaniach statystycznych używane jest również określenie „obcinania danych”.

⁸ $\delta_j=0$ – obserwacja cenzurowana, $\delta_j=1$ – obserwacja kompletna.

w badaniach medycznych nie jest tak oczywiste, zatem w artykule oparto się na estymatorach J.D. Kalbfleischa i R.L. Prentice'a niewymagających spełnienia tego założenia. Ich cechą charakterystyczną jest stała wartość w podprzedziałach $[t_{s-1}, t_s)$, $s=1, 2, \dots, j, \dots, r$, granice których wyznaczają momenty czasowe $0 < t_1 < t_2 < \dots < t_{s-1} < t_s < \dots < t_j < \dots < t_r < \dots$, w których występują zdarzenia niepożądane na jednostkach kohorty scharakteryzowanych wektorami wartości liczbowych cech $x(1), x(2), \dots, x(s-1), x(s), \dots, x(j), \dots, x(r), \dots$ odpowiednio. Zatem, jeśli $\xi(s) \equiv \xi(t_s), S^\wedge(j) \equiv S^\wedge(0, t_j)$ oraz $h^\wedge(s) \equiv h^\wedge(0, t_s)$, to estymatory $h^\wedge(s)$ oraz $S^\wedge(j)$ w sensie J.D. Kalbfleischa i R.L. Prentice'a mają następującą postać [Kalbfleisch, Prentice, 1973, s. 267–278]:

$$S^\wedge(j) = \prod_{s=1}^j \xi(s), \quad h^\wedge(s) = \frac{1 - \xi(s)}{t_s - t_{s-1}}, \quad \xi(s) = \left(1 - \frac{\exp[\beta \cdot x(s)]}{\sum_{i \in \Lambda(s)} \exp[\beta \cdot x(i)]} \right)^{\exp[\beta \cdot x(s)]} \quad (10)$$

IDENTYFIKACJA DETERMINANT RYZYKA W GRUPIE CHORYCH Z RAKIEM GŁOŚNI

W okresie 2002–2007 analizie poddano 122-osobową grupę pacjentów leczonych z powodu miejscowo zaawansowanego raka głośni przy wykorzystaniu radioterapii w Wojewódzkim Szpitalu Specjalistycznym w Rzeszowie (dane osobiste autorów oraz z Podkarpackiego Rejestru Chorób Nowotworowych). Rak głośni, czyli środkowego piętra krtani jest jednym z najczęstszych nowotworów anatomicznego regionu głowy i szyi. Jest to choroba rokująca stosunkowo dobrze, szczególnie w niskich stopniach zaawansowania. W jej leczeniu stosuje się dwie jednakowo skuteczne metody: leczenie operacyjne lub radioterapię. Główną przyczyną niepowodzeń leczenia są nawroty miejscowe lub regionalne. Oceniano czas przeżycia wolny od niepowodzeń leczenia T (ang. *Diseases free survival* – DFS) liczony od momentu zakończenia leczenia do wystąpienia wznowy miejscowej. Chorych różnicują zmienne objaśniające $X = \langle X_1, X_2, \dots, X_7 \rangle$ głównie o charakterze demograficznym oraz społeczno-ekonomicznym dostępne w dokumentacji lekarskiej i administracyjnej szpitala, których warianty przedstawia tabela 1. Za determinanty ryzyka uznane zostają te zmienne objaśniające, których wpływ na zaistnienie zdarzenia niepożądanego jest istotny.

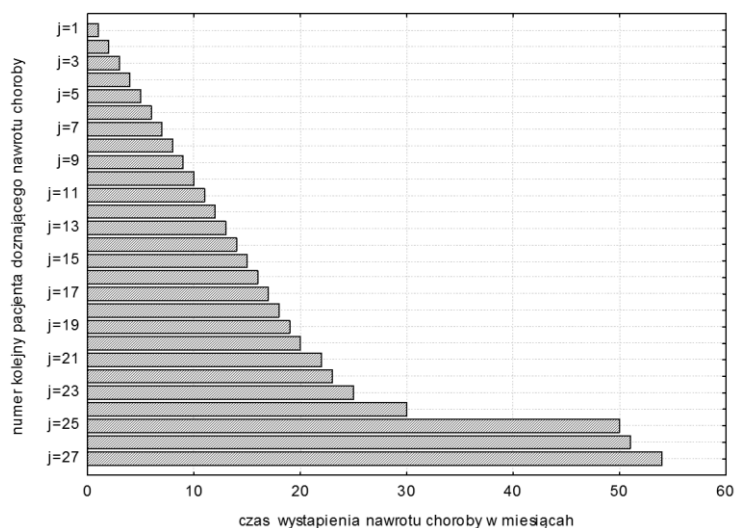
W badanej 122-osobowej kohorcie (grupie chorych) kobiety stanowiły 5,7% ogółu. Średnia wieku wynosiła 60,3 lat, przy odchyleniu standardowym 10,7 lat. Najmłodszy pacjent miał 36 lat, a najstarszy 83 lata. Wieś jako miejsce zamieszkania podało 66,4% osób. Dostęp do służby zdrowia jako zły oceniło 55,7% pacjentów, średni – 25,4%, a dobry 18,8%. W analizowanej grupie wykształcenie co najmniej średnie posiadało 22,1% ogółu badanych. Aktywnych zawodowo było 24,6% osób z ogółu, a pozostali to niepracujący (bezrobotni, emeryci i renciści). Wszyscy chorzy byli leczeni z powodu raka głośni w zaawansowaniu miej-

scowym, przy czym u 36,9% ogółu był to stopień najniższy T1 – a u pozostałych stopień zaawansowania wysoki – T2. W trakcie 5-letniej obserwacji odnotowano zajście 27 zdarzeń niepożądanych, czyli nawrotów miejscowych choroby. Obserwacje dla pozostałych pacjentów uznano jako cenzurowane. Czas wystąpienia niepowodzeń ilustruje rys. 1.

Tabela 1. 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 ∈ R ⁺
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	Mały – T1	x ₇ =0
		Duży – T2	x ₇ =1

Źródło: opracowanie własne oparte o dane kliniczne.



Rysunek 1. Czas wystąpienia nawrotu choroby liczony w miesiącach od momentu rozpoczęcia leczenia

Źródło: opracowanie własne.

W tabeli 2 wyszczególniono dostępne cechy charakteryzujące pacjentów doznających niepowodzenia w leczeniu w kolejności zgodnej z czasem wystąpienia wznowy. Natomiast czas wystąpienia zdarzenia niepożądanego, tj. nawrotu choroby liczony w miesiącach od momentu rozpoczęcia leczenia jest odczytywany z wykresu na rys. 1.

Tabela 2. Cechy charakteryzujące pacjentów uszeregowanych zgodnie z czasem wystąpienia u nich zdarzenia niepożądanego

Nr zdarzenia zgodny z rys. 1	X ₁ płeć	X ₂ wiek	X ₃ miejsce zamieszkania	X ₄ dostęp do służby zdrowia	X ₅ wykształcenie	X ₆ aktywność zawodowa	X ₇ zaawansowanie choroby
j=1	m	82	wieś	dobry	inne	nie	T2
j=2	m	63	wieś	zły	inne	nie	T2
j=3	m	62	wieś	zły	inne	nie	T2
j=4	m	65	wieś	dobry	inne	nie	T2
j=5	m	69	wieś	zły	inne	nie	T2
j=6	m	70	wieś	zły	inne	nie	T2
j=7	m	79	miasto	dobry	inne	nie	T2
j=8	m	72	wieś	zły	inne	nie	T2
j=9	m	54	wieś	zły	inne	nie	T2
j=10	m	59	miasto	dobry	inne	nie	T2
j=11	m	44	wieś	zły	inne	tak	T2
j=12	m	72	wieś	zły	inne	nie	T2
j=13	m	50	wieś	dobry	wyższe	nie	T2
j=14	m	71	wieś	zły	inne	nie	T2
j=15	m	69	wieś	zły	inne	nie	T1
j=16	m	60	wieś	zły	inne	nie	T2
j=17	m	47	miasto	dobry	inne	nie	T2
j=18	m	55	wieś	zły	inne	tak	T1
j=19	k	75	miasto	b. dobry	inne	nie	T2
j=20	m	71	miasto	dobry	wyższe	nie	T1
j=21	m	53	miasto	b. dobry	inne	nie	T2
j=22	m	53	miasto	b. dobry	inne	nie	T2
j=23	m	53	wieś	zły	inne	nie	T2
j=24	m	54	wieś	dobry	inne	nie	T2
j=25	m	45	wieś	zły	inne	nie	T1
j=26	m	64	miasto	zły	inne	nie	T2
j=27	m	57	wieś	zły	inne	tak	T1

Źródło: opracowanie własne w oparciu o dane kliniczne.

Teraz zgodnie ze wzorem (8) wyznacza się funkcję częściowej wiarygodności $L(\beta)$ będącej iloczynem 27 prawdopodobieństw warunkowych związanych z niepowodzeniami w leczeniu poszczególnych pacjentów kohorty (grupy). Roz-

wiązania układu równań $\text{grad}L(\beta)=0$, $\beta=\langle\beta_1,\beta_2,\dots,\beta_7\rangle$ wyznaczono przy użyciu pakietu *STATISTICA 6.0 PL* i otrzymano następujące wartości parametrów przyjętych do scharakteryzowania kohorty: $\beta_1=0,735$, $\beta_2=0,034$, $\beta_3=0,135$, $\beta_4=-0,054$, $\beta_5=1,404$, $\beta_6=-0,683$ oraz $\beta_7=0,987$. Metodą krokową wsteczną, szczegółowo opisaną w literaturze [Collet, 1994 lub Ostasiewicz (red.), 2000], dokonano optymalizacji modelu statystyką testową Walda odrzucając parametry nieistotne, a funkcjami częściowej wiarygodności i testem chi-kwadrat ustalono model optymalny. Najlepszy model otrzymano dla trzech zmiennych objaśniających: $X=\langle X_2,X_5,X_7\rangle$ przy optymalnych współczynnikach $\beta=\langle\beta_2=0,04;\beta_5=1,4;\beta_7=1,06\rangle$. Zatem ogólna postać funkcji ryzyka $h(x,t)$ dla opisywanego modelu optymalnego ma postać:

$$h(x,t)=h(0,t)\cdot\exp(0,04 X_2+1,40 X_5+1,06 X_7),$$

a pełne wyniki dotyczące testu Walda dla tego modelu zawiera tabela 3.

Tabela 3. Wyniki estymacji parametrów modelu

Zmienna	Ocena współczynnika β	Błąd średni szacunku współczynnika β	$\exp \beta$	Wartość statystyki Walda	Poziom p istotności 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, przy użyciu programu *STATISTICA 6.0*.

POSTAĆ TABELARYCZNA WYESTYMOWANYCH FUNKCJI PRZEŻYCIA I RYZYKA DLA WYBRANYCH PACJENTÓW

Analizy o konkretnych przebiegach funkcji przeżycia i ryzyka dla poszczególnych pacjentów są możliwe po wcześniejszym uzyskaniu wartości estymatorów $h^{\wedge}(t)$ i $S^{\wedge}(t)$ zgodnie ze wzorami (10). W tym celu całkowity czas obserwacji $[0,t_{r=27})=[0,60)$ podzielono na 27 podprzedziałów związanych z kolejnymi niepowodzeniami leczenia. W każdym z nich obliczono wartości $\xi(j)$, $S^{\wedge}(j)$ oraz $h^{\wedge}(j)$ dla $j=1,2,\dots,27$.

Oszacowane wartości funkcji przeżycia oraz ryzyka dla trzech przykładowych pacjentów $\omega_1, \omega_2, \omega_3$ o wektorach cech: $x(1)=\langle x_2(1)=80 \text{ lat}; x_5(1)=1 - \text{wykształcenia podstawowe}; x_7(1)=1 - \text{zaawansowanie choroby duże}\rangle$, $x(2)=\langle x_2(2)=60 \text{ lat}; x_5(2)=1 - \text{wykształcenia podstawowe}; x_7(2)=0 - \text{zaawansowanie choroby małe}\rangle$, $x(3)=\langle x_2(3)=60 \text{ lat}; x_5(3)=1 - \text{wykształcenia podstawowe}; x_7(3)=1 - \text{zaawansowanie choroby duże}\rangle$, dla kolejnych przedziałów czasowych $[t_{j-1},t_j)$, $j=1,2,\dots,27$ zawiera tabela 4.

Spostrzeżeniem z analizy tabeli 3 dotyczącej wartości funkcji przeżycia $S(x(\omega),t_j)$ określonego przedziału czasowego $[0,t_j)$ jest między innymi to, że pacjenci bez wykształcenia wyższego lub starsi wiekiem posiadają indywidualne funkcje przeżycia o znaczącym, monotonicznym spadku wartości w porównaniu

z funkcją bazową $S^{\wedge}(j)$ posiadającą praktycznie stałą, równą 0,99 wartość. Monotoniczny spadek jeszcze bardziej zwiększa się, gdy stopień zaawansowania choroby jest duży.

Tabela 4. Wyestymowane wartości funkcji ryzyka i funkcji przeżycia z uwzględnieniem cech charakteryzujących wybranych pacjentów

Przedział [t _{j-1} , t _j) w mie- siącach	S [∧] (j)	h [∧] (j)	Pacjent					
			ω ₁ , x(1)=<80;1;1>		ω ₂ , x(2)=<60;1;0>		ω ₃ , x(3)=<60;1;1>	
			S(x(1),t)	h(x(1),t)	S(x(2),t)	h(x(2),t)	S(x(3),t)	h(x(3),t)
[0-1)	0,99	9,1E-05	0,97	0,29	0,99	0,004	0,99	0,01
[1-2)	0,99	9,3E-05	0,95	0,29	0,99	0,004	0,98	0,01
[2-3)	0,99	9,4E-05	0,92	0,29	0,99	0,004	0,96	0,01
[3-4)	0,99	9,5E-05	0,90	0,30	0,98	0,004	0,95	0,01
[4-5)	0,99	9,9E-05	0,87	0,31	0,98	0,004	0,94	0,01
[5-6)	0,99	0,0001	0,85	0,32	0,97	0,004	0,93	0,01
[6-7)	0,99	0,0001	0,82	0,34	0,97	0,005	0,92	0,01
[7-8)	0,99	0,0001	0,80	0,35	0,96	0,005	0,90	0,01
[8-9)	0,99	0,0001	0,77	0,36	0,96	0,005	0,89	0,01
[9-10)	0,99	0,0001	0,74	0,39	0,95	0,005	0,87	0,01
[10-11)	0,99	0,0001	0,72	0,40	0,95	0,006	0,86	0,01
[11-12)	0,99	0,0001	0,69	0,41	0,94	0,006	0,85	0,02
[12-13)	0,99	0,0001	0,66	0,42	0,94	0,006	0,83	0,02
[13-14)	0,99	0,0001	0,64	0,44	0,93	0,006	0,82	0,02
[14-15)	0,99	0,0001	0,61	0,45	0,93	0,006	0,80	0,02
[15-16)	0,99	0,0001	0,59	0,45	0,92	0,006	0,79	0,02
[16-17)	0,99	0,0001	0,56	0,46	0,91	0,006	0,77	0,02
[17-18)	0,99	0,0001	0,54	0,46	0,91	0,006	0,76	0,02
[18-19)	0,99	0,0001	0,52	0,47	0,90	0,007	0,74	0,02
[19-20)	0,99	0,0001	0,50	0,48	0,90	0,007	0,73	0,02
[20-22)	0,99	7,9E-05	0,47	0,25	0,89	0,003	0,71	0,01
[22-23)	0,99	0,0002	0,45	0,51	0,88	0,007	0,70	0,02
[23-25)	0,99	8,2E-05	0,43	0,26	0,88	0,004	0,69	0,01

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

Natomiast intensywność hazardu (tj. funkcja ryzyka) ściśle związana z prawdopodobieństwem wystąpienia zdarzenia niepożądanego (zob. wzór (2) i (3)) po upływie roku w przedziale czasowym [12,13), tj. od 12 do 13 miesiąca od rozpoczętego leczenia w porównaniu z nikłym ryzykiem bazowym $h^{\wedge}(t=12)=0,0001$ charakterystycznym dla całej grupy leczonych jest dla pacjenta ω_1 mającego 80 lat, podstawowe wykształcenie i wysoki stopień zaawansowania choroby $4200=h(x(1),t=12)/h^{\wedge}(12)$ razy większa. Dla pacjenta ω_3 mającego 60 lat o podobnym wykształceniu i stopniu zaawansowania choroby funkcja ryzyka jest istotnie mniejsza od pacjenta ω_1 jednak i tak bardzo wysoka, bo 200 razy większa od

wzorca grupowego. Pacjent ω_2 mający też 60 lat i wykształcenie podstawowe, ale niski stopień zaawansowania choroby ma znacząco mniejsze ryzyko wznowy względem poprzedników, ale jest ono $60=h(x(2),t=12)/h^{(12)}$ razy większe od wzorca.

Zatem jest to potwierdzeniem faktu, że prawdopodobieństwo zaistnienia wznowy choroby, tj. zdarzenia niepożądanego pomiędzy 12 i 13 miesiącem od rozpoczęcia leczenia, istotnie wzrasta dla pacjentów starszych, szczególnie z wykształceniem podstawowym i dużym stopniem zaawansowania choroby.

WPLYW CZYNNIKÓW ISTOTNYCH NA RYZYKO WZGLĘDNE

Oszacowane wartości dla parametrów $\beta_2, \beta_5, \beta_7$ adekwatnych dla kohorty w modelu optymalnym pozwalają obliczyć, zgodnie ze wzorem (5) współczynniki ryzyka $HR_{j,s}=h(t,x(j))/h(t,x(s))$ dla pacjenta ω_j posiadającego cechy scharakteryzowane wektorem $x(j)=\langle x_2(j), x_5(j), x_7(j) \rangle$ w porównaniu do wzorca, którym może być każdy inny pacjent ω_s scharakteryzowany wektorem cech $x(s)=\langle x_2(s), x_5(s), x_7(s) \rangle$. Wartości względnego ryzyka dla różnych wariantów wartości cech X_5 i X_7 oraz różnicy wieku $\tau=x_2(j)-x_2(s)$ przedstawia tabela 5. Przyjęta w tej tabeli dyskretyzacja zmiennej τ co 5 lat upraszcza obliczenia i akcentuje możliwości poznawcze metody.

Analiza danych zawartych w tabeli wykazuje na to, że ryzyko niepowodzenia leczenia dla pacjenta w stopniu zaawansowania T2 jest 2,886-krotnie większe niż dla pacjenta w zaawansowaniu T1, przy takich samych wartościach pozostałych zmiennych. Osobę o wykształceniu podstawowym charakteryzuje 4,055-krotnie większe ryzyko nawrotu choroby w stosunku do osoby z wykształceniem wyższym, przy jednakowych wartościach pozostałych cech.

Możliwe są jeszcze inne warianty porównań opartych na wskaźniku ryzyka $HR_{j,s}$ jak na przykład ocena wpływu wykształcenia na ryzyko zaistnienia zdarzenia niepożądanego przy różnych stopniach zaawansowania choroby oraz przy nieujemnej różnicy wieku $\tau=x_2(j)-x_2(s) \in \{0,5,10,15,20,25,30\}$.

Pacjenci ω_j oraz ω_s o wektorach cech $x(j), x(s)$ odpowiednio mają tą samą szansę przeżycia, gdy współczynnik relatywnego ryzyka $HR_{j,s}=1$, a gdy $HR_{j,s}>1$ szanse przeżycia tego samego przedziału czasowego pacjenta ω_j są mniejsze niż pacjenta ω_s . Współczynnik relatywnego ryzyka $HR_{j,s}$ osiąga wartości większe od 1, gdy pacjent ω_j nie jest wykształcony, a znacząco większe od 1, gdy jest przy tym starszy od pacjenta ω_s i jeszcze bardziej większe, gdy pacjent ω_j ma chorobę o znacznym stopniu zaawansowania, tj. gdy $x_7(j)=1$. Ponadto, w przypadku, gdy choroba ma ten sam stopień zaawansowania, to lepiej wykształcony pacjent ω_j ma większe szanse przeżycia nawet wtedy, gdy od niewykształconego pacjenta ω_s jest starszy o 35 lat. Jeśli jego choroba jest bardziej zaawansowana, to $HR_{j,s}=1$

przy różnicy wieku większej od pięciu, ale mniejszej od dziesięciu lat, a przy mniej zaawansowanej $HR_{j,s} < 1$ nawet wtedy, gdy pacjent ω_j jest starszy od ω_s o więcej niż 40 lat.

Tabela 5. Wpływ czynników istotnych (determinant) na współczynnik ryzyka

HR _{j,s}	x _{7(j)} =x _{7(s)}			x _{7(j)} =1, x _{7(s)} =0			x _{7(j)} =0, x _{7(s)} =1		
	x _{5(j)} = x _{5(s)}	x _{5(j)} =1, x _{5(s)} =0	x _{5(j)} =0, x _{5(s)} =1	x _{5(j)} = x _{5(s)}	x _{5(j)} =1, x _{5(s)} =0	x _{5(j)} =0, x _{5(s)} =1	x _{5(j)} = x _{5(s)}	x _{5(j)} =1, x _{5(s)} =0	x _{5(j)} =0, x _{5(s)} =1
-5	0,819	3,320	0,202	2,363	9,583	0,583	0,284	1,150	0,070
0	1,000	4,055	0,247	2,886	11,705	0,712	0,346	1,405	0,085
5	1,221	4,953	0,301	3,525	14,296	0,869	0,423	1,716	0,104
10	1,492	6,050	0,368	4,306	17,462	1,062	0,517	2,096	0,127
15	1,822	7,389	0,449	5,259	21,328	1,297	0,631	2,560	0,156
20	2,226	9,025	0,549	6,424	26,050	1,584	0,771	3,127	0,190
25	2,718	11,023	0,670	7,846	31,817	1,935	0,942	3,819	0,232
30	3,320	13,464	0,819	9,583	38,861	2,363	1,150	4,665	0,284
35	4,055	16,445	1,000	11,705	47,465	2,886	1,405	5,697	0,346
40	4,953	20,086	1,221	14,296	57,974	3,525	1,716	6,959	0,423

Źródło: opracowanie własne na podstawie danych klinicznych z wykorzystaniem programu STATISTICA 0,5.

ZAKOŃCZENIE

Skuteczne zarządzanie ochroną zdrowia w społeczeństwie informacyjnym wymaga zastosowania technologii teleinformatycznych w identyfikacji czynników ryzyka wpływających na powstanie, przebieg i rokowania w leczeniu i zapobieganiu chorobom, a szczególnie chorobom rujnującym zdrowie (np. nowotworowym lub kardiologicznym). Dla jakości życia, zdrowie (rozumiane jako stan pełnego oraz dobrego samopoczucia fizycznego, psychicznego i społecznego, a nie tylko braku chorób lub niedomagań) jest najważniejszą wartością indywidualną człowieka i bezcennym dobrem społecznym. W zachowaniu oraz umacnianiu zdrowia oprócz stanu środowiska naturalnego, czynników genetycznych i działań opieki zdrowotnej istotną rolę odgrywa styl życia, któremu przypisuje się coraz większe znaczenie ze względu na wzrastające poczucie indywidualnej odpowiedzialności za stan zdrowia. Bieda, brak pracy, złe warunki pracy, niski poziom wykształcenia, złe warunki mieszkaniowe, stres, nieracjonalne odżywianie, palenie tytoniu itd. warunkują stan zdrowia określonych społeczności. Wnikliwa ocena tegoż zjawiska wymaga odpowiedniej metodologii badawczej.

We współczesnych badaniach medycznych, a zwłaszcza onkologicznych, dla jakości życia podstawowe znaczenie ma ocena czasu przeżycia pacjenta, czy też czasu do wznowy choroby zaleczonej, uwzględniającego istotne cechy o charak-

terze endogennym oraz egzogennym. Od momentu pokonania bariery obliczeniowej i przy łatwym dostępie do danych statystycznych w badaniach może być wykorzystywany model Coxa, opracowany teoretycznie na przełomie lat 70. i 80. XX wieku do problemów niezawodnościowych.

Przedstawiona w artykule idea konstruowania modelu proporcjonalnego ryzyka wraz z procedurą wyznaczenia ekonomiczno-społecznych cech determinujących jakość życia pacjenta leczonego onkologicznie w okresie 2002–2007 na raka głośni w Specjalistycznym Szpitalu Wojewódzkim im. Fryderyka Szopena w Rzeszowie jest tylko przykładem jego możliwości aplikacyjnych oraz interpretacyjnych. Zdaniem autorów, model Coxa z powodzeniem może być wykorzystany do wyznaczania determinant wpływających nie tylko na jakość życia, ale również na inne występujące w społeczeństwie informacyjnym zjawiska obciążone ryzykiem tj. dodatnim prawdopodobieństwem pojawiania się zdarzeń niepożądanych.

BIBLIOGRAFIA

- Biernacki M., 2009, *Dobrobyt* [w:] *Wybrane modele matematyczne w ekonomii – Prognozowanie i jakość*, red. J. Łyko, Wyd. Uniwersytetu Ekonomicznego we Wrocławiu, Wrocław.
- Collet D., 1994, *Modelling Survival Data in Medical Research*, Chapman&Hall.
- Cox D.R., 1972, *Regression Models and Life-Tables*, „Journal of the Royal Statistical Society”, 34.
- Hibner E., 2003, *Zarządzanie w systemie ochrony zdrowia*, Wyższa Szkoła Humanistyczno-Ekonomiczna, Łódź.
- Instrukcja: *Statistica Manual*, 1997, Wyd. StatSoft Polska, t. 1–3.
- Kalbfleisch J.D., Prentice R.L., 1973, *Marginal likelihoods based on Cox's regression and life model*, Biometrika, Vol. 60.
- Kalbfleisch J.D., Prentice R.L., 2002, *The Statistical Analysis of Failure Time Data*, Wiley-Interscience.
- Kaplan E.L., Meier P., 1958, *Nonparametric estimation from incomplete observations*, „Journal of the American Statistical Association”, Vol. 53, No. 282.
- Kleinbaum D.G., 1997, *Survival Analysis*, Springer.
- Kopociński B., 1973, *Zarys teorii odnowy i niezawodności*, PWN, Warszawa.
- Miller R.G., 1988, *Survival Analysis*, John Wiley & Sons.
- Ostasiewicz S. (red.), 2000, *Metody oceny i porządkowania ryzyka w ubezpieczeniach życiowych*, Wydawnictwo Akademii Ekonomicznej we Wrocławiu, Wrocław.
- Skrzypek J. (red.), 2007, *Uwarunkowania jakości życia w społeczeństwie informacyjnym*, Wydział Ekonomiczny Uniwersytetu Marii Curie-Skłodowskiej w Lublinie, Lublin.
- Stanisz A., 2000, *Przstępny kurs statystyki*, StatSoft Polska, t. 2.

Streszczenie

Nieodłączną cechą społeczeństwa informacyjnego jest dbałość o różnorodne aspekty jakości życia. Istnieje wiele metod, w tym statystycznych, służących wyznaczaniu i ocenie wpływu determinant (cech) warunkujących jakość życia.

W pracy rozpatrywano możliwość zastosowania w tym celu modelu proporcjonalnej intensywności Coxa. Użyto wariantu modelu opisującego zjawiska, w którym zdarzenia niepożądane występują w rozłącznych przedziałach czasowych. Używając estymatorów Kalbfleische'a-Prentice'a wyznaczono funkcje przeżycia $S(x,t)$ i ryzyka $h(x,t)$ dla badanych jednostek zróżnicowanych wektorem cech endo- i egzogennych.

Omaiwane funkcje szacowano na przykładzie wyników leczenia onkologicznego grupy chorych na raka głośni. Badano wpływ cech o charakterze demograficznym, społecznym i ekonomicznym bezpośrednio na rezultaty terapii, a pośrednio na jakość życia.

Słowa kluczowe: jakość życia, choroby onkologiczne, funkcje przeżycia i ryzyka, model Coxa

**Cox Proportional Hazards Model as a Statistical Method
for Identifying the Determinants of Quality of Life of Oncology Patients**

Summary

An integral feature of the information society is its interest in various aspects of quality of life. There are many methods, including statistical, for identifying and evaluating the impact of those factors (features) which influence quality of life.

The study considers the possibility of using the Cox model of proportional intensity for this purpose. Use was made of a variant of the model describing the phenomenon in which adverse events occur in disjointed time intervals. Using Kalbfleische'a-Prentice estimators the survival functions $S(x, t)$ and the risk of $h(x, t)$ were determined for the tested units differentiated by endogenous and exogenous features.

These functions were estimated based upon the results of treatment of a group of patients suffering from cancer. Research was carried out into the direct influence of demographic, social and economic characteristics on the results of the therapy and indirectly on the quality of life.

Keywords: quality of life, oncological diseases, survival and risk/hazard functions, Cox model
JEL: C00, D80, I14, I15, C510