

Praca doktorska zatytułowana „Modelowanie ryzyka zachorowalności z wykorzystaniem analizy przeżycia” jest poświęcona tematyce zastosowania metod analizy przeżycia w modelowaniu ryzyka zachorowalności. Rozprawa składa się ze wstępu, 5 rozdziałów i zakończenia. Praca podzielona jest na dwie części: teoretyczną i aplikacyjną. Część teoretyczna podejmuje tematykę porównania semiparametrycznych i parametrycznych modeli przeżycia, która nie została przedstawiona w sposób wystarczający w literaturze przedmiotu. Część aplikacyjna prezentuje przykład zastosowania rozpatrywanych metod porównania modeli przeżycia w analizie ryzyka zachorowalności dla wybranych chorób układu krążenia (zawał serca, udar mózgu, nadciśnienie tętnicze, podwyższony poziom cholesterolu).

### **Hipotezy badawcze:**

W kontekście motywacji do podjęcia tematu pracy oraz celów rozprawy sformułowano następujące hipotezy badawcze:

- (1) Semiparametryczne i parametryczne modele przeżycia stanowią użyteczne narzędzie w modelowaniu ryzyka zachorowalności umożliwiające identyfikację czynników determinujących ryzyko.
- (2) Metody oceny i porównania semiparametrycznych i parametrycznych modeli przeżycia pozwalają na wybór modelu najlepszego pod względem: (a) zdolności dyskryminacyjnej, (b) kalibracji, (c) ogólnej jakości dopasowania.
- (3) Typ schorzenia determinuje wybór modelu na potrzeby identyfikacji czynników ryzyka zachorowalności.

### **Metody badawcze:**

Badania przeprowadzone w pracy obejmują analizę danych wygenerowanych w sposób losowy oraz analizę danych rzeczywistych zbieranych w ramach badania SHARE (Badanie Zdrowia, Starzenia się i Przechodzenia na Emeryturę w Europie). Dane wygenerowane w sposób losowy zostały wykorzystane w celu oceny metod porównania semiparametrycznych i parametrycznych modeli przeżycia w ramach teoretycznej części rozprawy. Dane zbierane w ramach badania SHARE zostały wykorzystane w celu prezentacji przykładu zastosowania rozpatrywanych metod porównania semiparametrycznych i parametrycznych modeli przeżycia w analizie ryzyka zachorowalności na wybrane schorzenia układu krążenia w części aplikacyjnej rozprawy.

### **Streszczenie rozdziałów:**

Wstęp pracy prezentuje uzasadnienie wyboru tematu rozprawy, cele pracy w obszarze teoretycznym i aplikacyjnym oraz hipotezy badawcze. Podkreślono znaczenie identyfikacji czynników ryzyka zachorowalności, ze szczególnym uwzględnieniem wybranych schorzeń układu krążenia, które stanowią główną przyczynę zgonów w Polsce, Europie i na świecie od wielu lat. Spadek umieralności oraz płodności prowadzą do intensyfikacji procesów starzenia się ludności, stąd można oczekiwać, że schorzenia, dla których wiek jest jednym z istotnych czynników ryzyka będą stanowić coraz poważniejszy problem w przyszłości. Badania wskazują, że wczesna identyfikacja czynników ryzyka zachorowalności oraz podjęcie działań profilaktycznych w odpowiednim momencie mogą prowadzić do zredukowania rzeczywistych wskaźników zachorowalności. Wśród rozpatrywanych czynników ryzyka zachorowalności znajdują się zarówno charakterystyki medyczne, jak i socjo-demograficzne oraz behawioralne. Niniejsza praca przedstawia, w jaki sposób metody analizy przeżycia mogą być zastosowane do identyfikacji czynników, które w istotny sposób determinują ryzyko zachorowalności. Szczególną uwagę poświęcono problematyce porównywania semiparametrycznych i parametrycznych modeli przeżycia, która nie jest w sposób dostateczny omówiona w literaturze przedmiotu.

Pierwszy rozdział prezentuje przegląd teoretycznych koncepcji ryzyka oraz metod, które są najczęściej wykorzystywane w celu oceny i pomiaru ryzyka. W szczególności, uwzględniono trzy współczesne definicje ryzyka, zgodnie ze Słownikiem Języka Polskiego, który określa ryzyka jako: (1) szansę niepowodzenia określonego działania, jeżeli jego wyniki nie są z góry znane, (2) odwagę

konieczną do podjęcia działań uwzględnionych w punkcie (1), (3) prawdopodobieństwo wystąpienia szkody. Koncepcja ryzyka jest złożona i trudno sformułować jednoznaczną definicję tego pojęcia. W pracy przyjęto klasyfikację podejścia do definicji ryzyka zaprezentowaną przez Renna (1992). Na potrzeby pracy wybrano techniczną definicję ryzyka. Podejście techniczne pozwala na przeprowadzenie obiektywnego pomiaru ryzyka i wyrażenie go w sposób ilościowy, stąd wybór tego podejścia był uzasadniony w kontekście analizy ryzyka zachorowalności. Wykorzystanie metod analizy przeżycia umożliwia uwzględnienie nie tylko samego faktu wystąpienia zdarzenia (tu: zdiagnozowania schorzenia), ale też momentu w czasie, kiedy zdarzenie miało miejsce (tu: wiek respondenta).

Drugi rozdział prezentuje tło teoretyczne dotyczące semiparametrycznych i parametrycznych modeli przeżycia, z uwzględnieniem definicji i założeń modeli, estymacji i metod weryfikacji. W pracy wykorzystano model proporcjonalnych hazardów Coxa oraz pięć modeli parametrycznych: wykładniczy, Weibulla, log-logistyczny, log-normalny, gamma.

W trzecim rozdziale zaprezentowano proces oceny i porównania dwóch typów modeli przeżycia: semiparametrycznych i parametrycznych. Modele różnią się między sobą w zakresie przyjętych założeń oraz metody estymacji, stąd problematyka porównywania semiparametrycznych i parametrycznych modeli przeżycia staje się złożonym zagadnieniem. Metody porównywania modeli przeżycia rozpatrywane w pracy przedstawiono w podziale na trzy kategorie: (1) miary dyskryminacji (krzywa ROC i pole pod krzywą AUC, krzywa LIFT, indeks zgodności), (2) miary dyskryminacji (wykresy kalibracyjne, współczynnik kalibracji), (3) miary ogólnej jakości dopasowania modelu (wykresy reszt, kryteria informacyjne, standaryzowany współczynnik zmienności). Przeprowadzono ocenę rozpatrywanych metod porównywania modeli przeżycia w oparciu o analizę danych wygenerowanych w sposób losowy. Dla znanego rozkładu czasu przeżycia wyznaczono poszczególne miary umożliwiające porównanie modeli. Na podstawie uzyskanych wyników, oceniono wrażliwość rozpatrywanych miar porównania modeli na wybór poprawnego modelu oraz na fakt pominięcia ważnych zmiennych objaśniających w modelu. Sformułowano rekomendacje dotyczące wykorzystania metod porównywania modeli przeżycia w każdej z trzech rozpatrywanych kategorii.

Czwarty rozdział stanowi wstęp do części aplikacyjnej pracy, prezentując założenia oraz podejście do modelowania ryzyka zachorowalności dla wybranych schorzeń układu krążenia. Przedstawiono bazę danych wykorzystaną do analizy, uwzględniając ocenę jakości danych oraz wskazanie ograniczeń dostępnych danych z punktu widzenia analizy ryzyka zachorowalności. Zdefiniowano modele, zmienną oznaczającą zdarzenie (tu: diagnozę schorzenia) oraz zmienną czasową (tu: wiek respondenta w momencie diagnozy) oraz zaprezentowano schemat modelowania.

W piątym rozdziale zaprezentowano wyniki estymacji modeli oraz wybór modelu przeżycia na potrzeby analizy ryzyka zachorowalności dla wybranych chorób układu krążenia. W pierwszym etapie oszacowano model proporcjonalnych hazardów Coxa oraz pięciu modeli parametrycznych dla każdego schorzenia osobno. W drugim etapie przeprowadzono ocenę i porównanie modeli przeżycia ze względu na zdolność dyskryminacyjną, kalibrację oraz ogólną jakość dopasowania, na podstawie miar porównania modeli przeżycia przedstawionych w rozdziale trzecim.

## **Wnioski:**

Najważniejsze wnioski dotyczą części teoretycznej pracy i obejmują wyniki uzyskane na podstawie analizy danych wygenerowanych w sposób losowy. Otrzymane rezultaty wykazały dużą wrażliwość rozpatrywanych metod porównywania modeli na fakt pominięcia zmiennych objaśniających w modelu. Wniosek ten przemawia za przyjęciem pierwszej hipotezy badawczej – modele przeżycia stanowią użyteczne narzędzie umożliwiające identyfikację czynników ryzyka zachorowalności z uwagi na fakt, że pominięcie ważnych czynników ryzyka jest odzwierciedlone w obniżonych wartościach miar dyskryminacji, kalibracji i jakości dopasowania modeli. Z drugiej strony, uzyskane wyniki wykazały słabą wrażliwość na dobór poprawnego modelu. W większości przypadków jedynie model wykładniczy odznaczał się słabszą zdolnością dyskryminacyjną, kalibracją i jakością dopasowania w przypadku odstępstwa od założenia o stałości hazardu bazowego w czasie, podczas gdy pozostałe modele reprezentowały zbliżoną zdolność dyskryminacyjną, kalibrację i jakość dopasowania. Przy spełnieniu założenia o stałości hazardu bazowego w czasie stwierdzono niewielkie różnice pomiędzy poszczególnymi modelami. Powyższe wnioski pozwoliły stwierdzić częściową słuszność hipotezy drugiej: miary dyskryminacji, jakości dopasowania oraz kalibracji nie zawsze pozwalają na wybór najlepszego modelu przeżycia. Trzecią hipotezę badawczą weryfikowano

w oparciu o wyniki uzyskane na podstawie analizy danych zgromadzonych w ramach badania SHARE. Uzyskane wyniki przemawiają za przyjęciem trzeciej hipotezy: w przypadkach, kiedy były podstawy do wyboru jednego z modeli, dla różnych schorzeń różne modele wybierano jako najlepsze pod względem zdolności dyskryminacyjnej, kalibracji i jakości dopasowania.

Podsumowując, analizy przeprowadzone w pracy pozwoliły stwierdzić, że modele przeżycia stanowią użyteczne narzędzie w analizie ryzyka zachorowalności. Poprawny dobór modelu okazał się nie być tak istotnym z punktu widzenia zdolności dyskryminacyjnej, kalibracji i jakości dopasowania (z wyjątkiem przypadku odstępstwa od założenia o stałości hazardu bazowego w czasie) jak fakt włączenia wszystkich istotnych zmiennych objaśniających do modelu. Wyniki otrzymane na podstawie przeprowadzonych analiz mogą stanowić punkt wyjścia do dalszych badań. W szczególności, istotne są następujące kwestie:

- w jakim stopniu skuteczna identyfikacja czynników ryzyka zachorowalności może przyczynić się do obniżenia rzeczywistej, obserwowanej zachorowalności lub śmiertelności (np. poprzez działania profilaktyczne, zmianę diety lub stylu życia)?
- czy i w jakim stopniu wczesna identyfikacja czynników determinujących ryzyko wystąpienia danej choroby może wpłynąć na obniżenie łącznych kosztów opieki medycznej (z uwzględnieniem kosztów poniesionych na działania zmierzające do zidentyfikowania takich charakterystyk oraz wszelkich działań profilaktycznych, a także ewentualnych kosztów leczenia, których na skutek efektywnej identyfikacji czynników ryzyka można uniknąć)?

Odpowiedzi na powyższe pytania wymagają dalszych badań uwzględniających analizę danych medycznych i ekonomicznych, niemniej na podstawie dotychczasowych badań można oczekiwać, że wczesna identyfikacja czynników ryzyka zachorowalności oraz podjęcie działań profilaktycznych będzie skutkować obniżeniem faktycznych wskaźników zachorowalności oraz łącznych kosztów opieki zdrowotnej (uwzględniających koszty działań profilaktycznych oraz potencjalne koszty leczenia, których można uniknąć).