

**Н. В. Ковтун,**

доктор економічних наук, професор,  
кафедра статистики та демографії,  
Київський національний університет імені Тараса Шевченка,  
E-mail: kovtun\_natali@ukr.net;

**І. М. Мотузюк,**

кандидат медичних наук, доцент,  
кафедра онкології,  
Національний медичний університет імені О. О. Богомольця,  
E-mail: 180978igor@gmail.com;

**Р. О. Ганжа,**

аспірант,  
Київський національний університет імені Тараса Шевченка,  
E-mail: ganzha.roman.alex@gmail.com

### Використання регресії Кокса для прогнозування виживаності жінок з множинними злоякісними новоутвореннями

На підставі попереднього аналізу здійснюється моделювання виживаності жінок, хворих на рак грудної залози (РГЗ), з метою пошуку найбільш суттєвих чинників, які не випадково впливають на імовірність дожиття жінок з множинними злоякісними новоутвореннями. Побудовано низку моделей виживаності – загальні й окремо для синхронних та метакронних процесів перебігу раку. Показано, що перебіг захворювання у пацієнок на синхронний рак є більш агресивним та прогностично несприятливим. Доведено, що принципово відрізняються фактори формування виживаності пацієнок у віці до і після 30 років на момент аварії на ЧАЕС. Вказується на можливість подальшого використання отриманих моделей для прогнозування виживаності жінок досліджуваних груп.

**Ключові слова:** регресія Кокса, прямий метод, аналіз виживаності, рак грудної залози, множинні первинні злоякісні новоутворення.

**Формулювання проблеми.** В останні роки спостерігається зростання частоти виникнення множинних первинних злоякісних новоутворень, коли дві або більше непов'язані первинні злоякісні пухлини, що походять з різних органів, з'являються в організмі одночасно або одна за одною. Так, в Україні інтервал між першим і другим діагнозом раку репродуктивної системи скоротився майже у шість разів з 11 до 2 років, а імовірність прожити наступні 3 роки після 8,5 років з моменту встановлення першого діагнозу зменшилася з 0,959 до 0,562.

Моделювання характеру дожиття онкохворих жінок потребувало попереднього оцінювання їх виживаності з урахуванням наявності у пацієнок синхронних (SMPMN) і метакронних (MMPMN) процесів перебігу раку. За даними проведеного вибіркового дослідження, результати якого представлені у [10], імовірність прожити ще 125 місяців у пацієнтів на SMPMN становила 0,73, тоді як у пацієнтів на склала 0,92 ( $p < 0,01$ ). Тобто перебіг захворювання у пацієнтів на синхронний рак є більш агресивним та прогностично несприятливим. До того ж доведено, що принципово відрізняються фактори формування виживаності па-

цієнок у віці до і після 30 років на момент аварії на ЧАЕС.

Отже, постало питання щодо можливості моделювання виживаності жінок, хворих на рак грудної залози, на основі пошуку найбільш суттєвих чинників, які не випадково впливають на імовірність дожиття жінок з множинними злоякісними новоутвореннями з метою подальшого використання цих моделей для прогнозування виживаності на певному ознаковому просторі.

**Аналіз останніх досліджень і публікацій.** Як свідчать міжнародні наукові дослідження, множинні первинні злоякісні новоутворення належать до розвитку злоякісного ураження *de novo*, їх характеристики повністю відрізняються від первинної пухлини [1]. Рак грудної залози складає 30–50% усіх других первинних злоякісних новоутворень у жінок, які мали первинний рак грудної залози (РГЗ). Вони мають у 2–6 разів вищий ризик виникнення нового первинного раку у контрлатеральній грудній залозі впродовж життя [2–4], що складає 0,3–1,0% щороку [3; 5]. У літературі існує невизначеність щодо прогнозу і виживаності хворих на метакронні онкологічні процеси порівняно з первинним одностороннім раком грудної залози. З огляду на це в роботі досліджуються закономірності виникнення метакронних процесів у

пацієнтів з пухлинами грудної залози, оцінюються їх наслідки та здійснюється прогноз.

Розвиток множинних процесів може бути пов'язаний з радіаційним забрудненням унаслідок аварії на ЧАЕС. Так, виявлено надлишкові радіаційно зумовлені випадки РГЗ серед опромінених жінок після атомного бомбардування Хіросіми та Нагасакі [1]. Розвиток цієї локалізації раку залежав від віку зазначеної групи населення на момент опромінення. Вища частота виникнення пухлин спостерігалася серед осіб, опромінених у віці 10–19 років. При опроміненні у віці 20 років і старше захворюваність на РГЗ була менша. Середній латентний період, після якого виникав РГЗ, не залежав від дози і складав приблизно 18 років. У світовій літературі визначено, що захворюваність на РГЗ серед жінок, опромінених у віці 10–29 років, вища ніж у віці старше 30 років [2].

Мета статті – моделювання виживаності жінок, хворих на рак грудної залози, на основі пошуку найбільш суттєвих чинників для подальшого використання цих моделей у прогнозуванні виживаності у визначеному факторному середовищі, що не випадково впливає на імовірність дожиття жінок з множинними злоякісними новоутвореннями.

**Методологія.** Для моделювання виживаності жінок, хворих на рак грудної залози з множинними злоякісними новоутвореннями було використано регресію Кокса [4]. Основа перевага цієї моделі полягає в тому, що вона дає можливість працювати з категоріальними та цензурованими даними. Загалом модель можна представити в такому вигляді:

$$h(t) = h_0(t) \cdot \exp(b_1x_1 + b_2x_2 + \dots + b_kx_k),$$

де  $h(t)$  – функція ризику, визначена набором  $k$  факторів  $x_1, x_2, \dots, x_k$ ;  $t$  – час виживання;  $b_1, b_2, \dots, b_k$  – коефіцієнти, що визначають вплив  $x_1, x_2, \dots, x_k$  відповідно;  $h_0$  – базовий ризик.

З метою поглибленого аналізу імовірності виникнення рецидивів та метастаз за різними факторами було використано логістичну регресію за технікою оптимізації Fisher's scoring.

Для загальної характеристики виживаності до основи вибірки було включено 2032 пацієнтки, які отримали спеціальне лікування у відділенні пухлин грудної залози та її реконструктивної хірургії Національного інституту раку з 2008 року по грудень 2015 року в межах рандомізованого контрольованого відкритого дослідження з вивчення критеріїв об'єктивізації вибору обсягу оперативного втручання у хворих на РГЗ. Окрему групу складала 195 (9,6%) пацієнток із множинними злоякісними новоутвореннями. Пацієнток було поділено на 4 групи: 1) із синхронним раком грудних залоз (SBC); 2) із синхронним раком грудної залози та іншої локалізації (SBCO); 3) із метакронним раком грудних залоз (MBC); 4) із

метакронним раком грудної залози та іншої локалізації (MBCO).

З числа відібраних 195 пацієнток у 107 осіб (54,9%) відбувалися синхронні процеси. Розподіл синхронних процесів за локалізацією був таким: 56% – грудна залоза і 44% – комбінація грудної залози з іншими локалізаціями. На частку метакронних процесів припадало 45% пацієнток, з яких 46,6% мали локалізацію у грудній залозі, а решта – це комбінація нозологій.

Для моделювання після поглибленого аналізу було прийнято рішення будувати регресію Кокса методом forward. Його сутність полягає в тому, що включення факторів у модель починається з меншої кількості ефектів і крок за кроком прямує до більшої кількості з можливістю тільки додавати ефекти до моделі на кожній ітерації, на відміну від методу stepwise, де передбачається додавання та видалення ефектів з моделі на кожній ітерації. Щодо backward методу, то першим кроком він обробляє всі ефекти і далі починає крок за кроком видаляти ефекти з моделі. Якщо ефект видаляється на якомусь кроці, то він більше не додається до моделі на будь-якому іншому кроці. Регресія Кокса розрахована на роботу з цензурованими даними.

Стандартна модель даних була побудована в Excel, а аналітична модель даних – у програмному середовищі SAS 9.4. Усі розрахунки були також зроблені у програмному середовищі SAS 9.4.

## Результати

### 1. Статистична оцінка виживаності жінок з множинними злоякісними новоутвореннями.

Задля реалізації мети було побудовано низку моделей виживаності – загальні й окремо для синхронних та метакронних процесів. Для побудови моделей було визначено факторне середовище, до якого увійшли такі змінні:

2. Тривалість періоду дожиття (змінна DY) – час у днях з моменту встановлення першого діагнозу до контрольної дати (на яку визначено статус пацієнта: живий або помер).

3. Вид множинного процесу (змінна ChromeGR) – синхронний або метакронний рак.

4. Момент встановлення першого діагнозу (змінна StartDT1GR) – до і після 2008 року.

5. Наявність рецидивів та метастазів упродовж досліджуваного періоду (змінна RecGR).

6. Вікова група на момент аварії на ЧАЕС: до 30 років і понад 30 років (змінна AgeGR).

7. Вік пацієнта на контрольну дату (змінна Age). Середній вік хворих на MPMNs становив 46,6 року. Розподіл хворих за віком такий: жінки репродуктивного віку – 72 (37%), жінки у постменопаузі – 123 (63%).

8. Стадія захворювання (змінна Stage) на момент першого діагнозу раку грудної залози.

До вибірки включили хворих на РГЗ зі стадіями I – IIIС. Розподіл за стадіями такий: процес  $T_{1-2}N_0M_0$  спостерігався у 114 хворих (58,5%),  $T_{1-2}N_1M_0$  – у 35 хворих 17,9% і  $T_{1-4}N_{1-3}M_0$  – у 46 хворих (23,6%). Зазначений розподіл демонструє агресивність цих варіантів РГЗ.

9. Комбінація лікування (змінна OperateGR1): оперативне втручання; хіміотерапія; опромінювання; гормональна терапія.

10. Наявність хіміотерапії у лікуванні (змінна OperateGR2).

11. Вид операції: мастектомія та органозберігаюча (змінна OperationGR).

12. Методика проведення операції: стандартна та сучасна (змінна DoctorGR).

13. Перша локалізація (LocalGR) – грудна залоза, репродуктивні органи, інші локалізації.

За результуючу змінну було визначено тривалість періоду дожиття. Цензурована змінна – статус пацієнта на контрольну дату.

З метою оцінювання ризиків виживаності було побудовано серію моделей: загальну, для всієї сукупності із включенням обох видів множинного процесу (синхронні та метасинхронні), а також часткові моделі, тобто окремо для синхронних та метасинхронних процесів (табл. 1, власні розрахунки на основі статистики Національного Інституту раку). Сутність моделі полягає в тому, що за результатами перевірки глобального тесту нульові гіпотези ( $Beta[i] = 0$ ) відхиляються на користь альтернативних ( $Beta[i] \neq 0$ ) а це говорить про те, що кінцеві моделі мають сенс і можуть прогнозувати виживаність. Так, гіпотези про відсутність істотних невідповідних чинників, що визначають виживаність, так звані глобальні гіпотези, перевірялися на істотність і були відхилені на користь альтернативних, а отже, кінцеві моделі мають сенс і можуть бути використані для прогнозування виживаності жінок, хворих на рак грудної залози.

Таблиця 1

Загальний вигляд рівняння регресії Кокса (коефіцієнти регресії) для трьох моделей

Ефект	Оцінки параметрів		
	Загальна модель	Синхронні процеси	Метасинхронні процеси
ChromeGR	2,53860*	0,87098	-0,66822
RecGR	2,30869*	18,87591	2,14296*
OperateGR1 1 (4)	-1,81157	14,62330	-2,28599
OperateGR1 2 (4)	0,05154	-0,22191	0,91613
OperateGR1 3 (4)	-0,14100	-0,49428	-0,14995
OperateGR2	-3,54445*	14,73657	-4,46091*
OperationGR	-0,26688	-0,91125	-0,72340
DoctorGR	-0,77000	-1,39052	-0,34877
StageGR 1 (3)	-2,45215*	-18,25191	-2,55378*
StageGR 2 (3)	-0,48085	-0,24833	-1,81311*
AgeGR	-0,17462	-0,54278	-0,69405
Age	-0,00977	-0,00578	-0,09855

\* достовірні коефіцієнти

За результатами перевірки глобального тесту доведено можливість прогнозування виживаності як для загальної моделі, так і для моделі, побудованої для метасинхронних процесів на рівні істотності 0,05 (табл. 2, власні розрахунки на основі статистики Національного Інституту раку). Щодо

моделі для синхронних процесів, то для доведення її адекватності не було достатньо кількості цензурованих спостережень. Саме тому не вдалося довести можливість її використання для прогнозування виживаності хворих на синхронний рак.

Таблиця 2

Результати глобального тесту для перевірки нульової гіпотези  $Beta[i] = 0$  для трьох моделей

Тест	Pr > ChiSq		
	Загальна модель	Синхронні процеси	Метасинхронні процеси
Likelihood Ratio	<,0001	0,2201	0,1129
Score	<,0001	0,5993	<,0001
Wald	0,0275	0,9987	0,2108

Попередній висновок підтверджується і даними табл. 3 (власні розрахунки на основі статистики Національного Інституту раку), з якої можна побачити, що найбільш суттєвими факторами прогнозування виживаності для загальної вибірки на рівні істотності 0,05 виявилися: вид множинного процесу (наявність синхронних або метасинхронних процесів), наявність ре-

цидивів, вид та комбінація лікування, стадія захворювання. Для моделі, що побудована для метасинхронних процесів, важливими факторами є: вид та комбінація лікування, наявність рецидивів, стадія захворювання. Щодо моделі для синхронних процесів, то жоден з ефектів не виявився істотним.

Результати Type 3 Test of Forward Selection

Ефект	Pr > ChiSq		
	Загальна модель	Синхронні процеси	Метахронні процеси
ChromeGR	0,0010*	0,4591	0,4809
RecGR	0,0320*	0,9969	0,1253**
OperateGR1	0,7291	0,9809	0,2893
OperateGR2	0,0013*	0,9991	0,0034*
OperationGR	0,6566	0,3912	0,5111
DoctorGR	0,2790	0,3491	0,7638
StageGR	0,0950*	0,9625	0,1691**
AgeGR	0,8514	0,6671	0,6897
Age	0,8482	0,9411	0,3089

\*) достовірні розбіжності; \*\*) розбіжності, на які слід звернути увагу.

Для оцінювання ризиків, пов'язаних з виживаністю, в рамках регресійного аналізу було використано Hazard Ratio, результати обчислення

якого представлені в табл. 4 (власні розрахунки на основі статистики Національного Інституту раку).

Таблиця 4

Результати розрахунку Hazard Ratio для оцінювання ризиків виживаності для різних моделей

Ефект	Hazard Ratio			Ранг за впливом на ризик
	Загальна модель	Синхронні процеси	Метахронні процеси	
ChromeGR	12,662	2,389	0,513	2
RecGR	10,061	1,5765E8	8,525	4
OperateGR1 1 (4)	0,163 (6,1)	2242946	0,102	5
OperateGR1 2 (4)	1,053	0,801	2,500	11
OperateGR1 3 (4)	0,868 (1,15)	0,610	0,861	10
OperateGR2	0,029 (34,5)	2511958	0,012	1
OperationGR	0,766 (1,3)	0,402	0,485	8
DoctorGR	0,463 (2,16)	0,249	0,706	6
StageGR 1 (3)	0,086 (11,6)	0,000	0,078	3
StageGR 2 (3)	0,618 (1,6)	0,780	0,163	7
AgeGR	0,840 (1,2)	0,581	0,500	9
Age	0,990 (1,01)	0,994	0,906	12

Результати розрахунків дають підстави для попередніх висновків:

1. Ризики виживаності у синхронного раку у 12,7 раза вищі, ніж у метахронного, що підтверджує попередню гіпотезу про набагато більшу агресивність синхронних процесів. Причому у синхронного раку з різними локалізаціями ризики більше у 2,4 раза, а у метахронного раку – практично удвічі.

2. Слід зазначити, що для множинних процесів істотне значення має проведення хіміотерапії у комбінації з іншими видами лікування (такими як хірургічне, променеве та медикаментозне), яка в 34,5 раза підвищує виживаність пацієнтів. Отже, за впливовістю на виживаність хіміотерапія посідає перше місце.

3. Третє місце за впливовістю на виживаність належить фактору, що характеризує наявність рецидивів і метастазів. Так, ризики виживаності за умов їх наявності в 10 разів вищі. Причому для синхронних процесів вони мають більші наслідки, ніж для метахронних, що дає підстави для висунення гіпотези

про високу імовірність прояву рецидивів і метастазів саме для синхронного раку.

4. Ризик виживаності для третьої стадії раку в 11,6 раза вищий, ніж для першої, і в 1,6 раза – ніж для другої. Отже, стадія захворювання також чинить не випадковий суттєвий вплив, що має бути враховано при прогнозуванні виживаності.

5. **Фактори оцінки імовірності рецидивів і метастазів у жінок з множинними злоякісними новоутвореннями.** Як було доведено раніше, рецидиви та метастази суттєво впливають на виживаність. Тому з метою поглиблення аналізу було досліджено розбіжності в імовірності виникнення рецидивів за різними параметрами, а саме: за наявністю хіміотерапії та без неї; за комбінацією лікування; за різних операцій для синхронних та метахронних процесів; за віком. Аналіз було здійснено з використанням процедури парної логістичної регресії, результати якої представлені в табл. 5 (власні розрахунки на основі статистики Національного Інституту раку).

Результати логістичної регресії імовірності виникнення рецидивів за різними факторами

Характеристика \ Ефект	OperateGR2	ChromeGR	OperateGR1	OperationGR	DoctorGR
Intercept	-0,9938*	0,0341	-3,2057	-0,0869	-0,1956
Coef, Regr,	1,3733**	-0,2404	-10,9; 3,09; 3,09	0,5512*	0,4707*
Likelihood Ratio	<,0001	0,0961	<,0001	0,0002	0,0036
Score	<,0001	0,0965	<,0001	0,0002	0,0038
Wald	<,0001	0,0973	0,0644	0,0003	0,0043
Type 3 Аналіз ефектів	<,0001	0,0973	0,0644	0,0003	0,0043
Odds Ratio	15,589	0,618	<0,001	3,011	2,564
Gamma	0,879	0,236	0,571	0,501	0,439

\* Pr > ChiSq = <0.01; \*\* Pr > ChiSq = <0.0001.

За результатами обчислень можна зробити додаткові цікаві висновки. Так, незважаючи на те, що хіміотерапія при множинних процесах суттєво знижує ризики виживаності, проте вона є водночас і визначальним фактором появи рецидивів та метастазів: шанси їх виникнення зростають в 15,6 раза за наявності хіміотерапії в комбінованому лікуванні. Зв'язок є не тільки істотним, а і достатньо тісним (0,879). Отже, при аналізі прямого впливу ми маємо нібито позитивний вплив хіміотерапії на виживаність, однак якщо включити фактор рецидивів, який підвищує ризики виживаності, то загалом можна зробити висновок про загальний негативний вплив хіміотерапевтичних процедур. Аналогічну ситуацію ми маємо і з комбінованим лікуванням.

Щодо сучасної методики проведення операції, то за результатами моделювання доведено, що вона порівняно з класичною в 2,6 раза зменшує ризик виникнення рецидивів і метастазів, а отже, позитивно впливає на імовірність дожиття і знижує ризик смертності. Продовжуючи тему операцій та методик їх проведення, однозначно можна стверджувати, що органозберігаючі операції при множинних онкологічних процесах утрічі знижують ризик виникнення рецидивів та метастазів, а це, своєю чергою, також підвищує імовірність дожиття. Вікова ознака виявилася несуттєвою з погляду впливу на імовірність появи рецидивів та метастазів.

**Висновки.** Глобальна гіпотеза про відсутність істотних невивадкових чинників, що визначають виживаність жінок з множинними онкологічними процесами, була перевірена на істотність. У результаті цього доведено, що розроблена модель регресії Кокса для загальної вибірки (195 жінок) має сенс і може бути використана для прогнозування виживаності жінок, хворих на рак грудної залози, за наявністю у них синхронних або метакронних процесів. За результатами моделювання виявлено найбільш істотні фактори, які невивадково впливають на виживаність. Такими факторами є: вид множинного процесу (наявність синхронних або метакронних

процесів); наявність рецидивів та / або метастазів; вид та комбінація лікування; стадія захворювання.

Синхронний рак характеризується більшою агресивністю і практично у 13 разів знижує виживаність порівняно з метакронним раком. Незважаючи на те, що хіміотерапія значно підвищує виживаність пацієнтів, вона одночасно провокує (супроводжує) появу рецидивів та метастазів, які в 16 разів частіше виникають за її наявності. Це дає підстави висунути гіпотезу про опосередкований вплив хіміотерапії на виживаність, який необхідно розглядати в контексті можливості її негативного впливу на появу рецидивів та метастазів, які, своєю чергою знижують виживаність у 10 разів. Встановлення цього факту, на нашу думку, ставить питання про необхідність додаткового поглибленого аналізу. Доведено істотну розбіжність у виживаності жінок з першою і третьою стадіями раку: шанси вижити за першої стадії практично в 12 разів вищі і це є логічним. Водночас різниця у виживаності жінок з другою і третьою стадіями незначна і складає тільки 1,6 раза.

Сучасна методика проведення операції порівняно зі стандартною в 2,6 раза зменшує ризик виникнення рецидивів і метастазів, а органозберігаючі операції при множинних онкологічних процесах – у три рази. Це дає підстави стверджувати, що вони позитивно впливають на імовірність дожиття і знижують ризик смертності.

Щодо часткових моделей, побудованих окремо для жінок, у яких спостерігається синхронний процес, і жінок з метакронним процесом, то для оцінювання можливих розбіжностей за факторами впливу на виживаність і покращення прогностичних властивостей моделей необхідно збільшення обсягу вибірки. А це, своєю чергою, потребує проведення додаткового дослідження, в ході якого будуть зібрані необхідні додаткові дані.

Цікавим є питання максимізації статистичних критеріїв – 2 LOG L, AIC, SBC, що показують адекватність підбору моделі, та коригування вищезгаданих методів підбору моделі, але це вже інше завдання, яке може бути темою окремого дослідження.



## References

1. Bongers, M. L., de Ruyscher, D., Oberije, C., Lambin, P., Uyl-de Groot C. A., & Coupe, V. M. H. (2016). Multistate statistical modeling: a tool to build a lung cancer microsimulation model that includes parameter uncertainty and patient heterogeneity. *Medical decision making*, 36 (1), 86–100. Retrieved from <https://doi.org/10.1177/0272989X15574500>
2. Fedorenko, Z. P., Michailovich, Yo. Yo., Goulak, L. O., Gorokh, Ye. L., Ryzhov, A. Yu., Soumkina, O. V., et al. (2016). *Cancer in Ukraine, 2014–2015. Incidence, mortality, activities of oncological service*. O. O. Kolesnik (Ed.). Bulletin of National Cancer Registry of Ukraine. № 17. Retrieved March 23, 2017, from [http://www.ncru.inf.ua/publications/BULL\\_17/index\\_e.htm](http://www.ncru.inf.ua/publications/BULL_17/index_e.htm)
3. Cancer fact sheets. Cancer today. International Agency of Research Cancer. *gco.iarc.fr*. Retrieved March 23, 2017, from <http://gco.iarc.fr/today/fact-sheets-cancers>
4. Collett, D. (2015). *Modelling Survival Data in Medical Research* (3rd ed.). Boca Raton: Chapman & Hall / CRC.
5. Heron, D. E., Komarnicky, L. T., Hyslop, T., Schwartz, G. F., & Mansfield, C. M. (2000). Bilateral breast carcinoma. *Cancer*, 88 (12), 2739–2750.
6. Jobsen, J., Palen, J. D., Ong, F., & Meerwaldt, J. (2003). Synchronous, bilateral breast cancer: prognostic value and incidence. *The Breast*, 12 (2), 83–88.
7. Lee, M. M., Heimann, R., Powers, C., Weichselbaum, R. R., & Chen, L. M. (1999). Efficacy of Breast Conservation Therapy in Early Stage Bilateral Breast Cancer. *The Breast Journal*, 5 (1), 36–41. doi:10.1046/j.1524-4741.1999.005001036.x
8. Levi, F., Randimbison, L., Te, V., & Vecchia, C. L. (2003). Prognosis of bilateral synchronous breast cancer in Vaud, Switzerland. *The Breast*, 12 (2), 89–91.
9. Merimsky, O., Kollender, Y., Issakov, J., Bickels, J., Flusser, G., Gutman, M., et al. (2001). Multiple primary malignancies in association with soft tissue sarcomas. *Cancer*, 91 (7), 1363–1371.
10. Kovtun, N., Palian, Z., & Motuziuk, I. (2017). Statistical analysis of trends and factors in cancer incidence of Ukrainian women in reproductive age. *The 11th Professor Aleksander Zelias International Conference on Modelling and Forecasting of Socio-economic Phenomena. Conference Proceedings. Cracow: Foundation of the Cracow University of Economics*, pp. 173–182.
11. Xu, L., & Gu, K. (2014). Clinical retrospective analysis of cases with multiple primary malignant neoplasms. *Genetics and Molecular Research*, 13 (4), 9271–9284.

**Н. В. Ковтун,**

доктор економічних наук, професор,  
кафедра статистики і демографії,  
Київський національний університет імені Тараса Шевченка;

**И. М. Мотузок,**

кандидат медичинських наук, доцент,  
кафедра онкології,  
Національний медичинський університет імені А. А. Богомольця;

**Р. А. Ганжа,**

аспірант,  
Київський національний університет імені Тараса Шевченка

## Использование регрессии Кокса для прогнозирования выживаемости женщин с множественными злокачественными новообразованиями

На основании предварительного анализа осуществляется моделирование выживаемости женщин, больных раком молочной железы (РГЗ), с целью поиска наиболее существенных факторов, которые случайно влияют на вероятность дожития женщин с множественными злокачественными новообразованиями. Построен ряд моделей выживаемости – общие и отдельно для синхронных и метакронного процессов течения рака. Показано, что течение заболевания у пациенток на синхронный рак является более агрессивным и прогностически неблагоприятным. Доказано, что принципиально отличаются факторы формирования выживаемости пациенток в возрасте до и после 30 лет на момент аварии на ЧАЭС. Указывается на возможность дальнейшего использования полученных моделей для прогнозирования выживаемости женщин исследуемых групп.

**Ключевые слова:** регрессия Кокса, прямой метод, анализ выживаемости, рак грудной железы, множественные первичные злокачественные новообразования.

**N. V. Kovtun,**

*DSc in Economics, Professor,  
Department of Statistics and Demography,  
Taras Shevchenko National University of Kyiv;*

**I. M. Motuziuk,**

*PhD in Medicine, Associate Professor,  
Department of oncology,  
O. O. Bogomolets National Medical University;*

**R. O. Ganzha,**

*Postgraduate Student,  
Taras Shevchenko National University of Kyiv*

## **Using Cox Regression to Forecast of Survival of Women with Multiple Malignant Neoplasms**

Recently, an increase in the incidence of multiple primary malignant neoplasms has been observed, specifically, when two or more unrelated tumors originate from different organs and appear in the body simultaneously or sequentially, one after another. During past few years, the interval between the first and second reproductive cancer diagnosis has decreased in 6 times – from 11 to just 2 years while probability of surviving the next 3 years after 8.5 years past initial diagnosis has decreased from 0.995 to 0.562. Using performed analysis, this paper provides details of survival modelling for women with breast cancer with the aim to find the most significant factors affecting the likelihood of survival not by chance alone. The data used for research were obtained from Ukrainian National Institute of Cancer covering 1981–2017 period.

The modelling was performed using Cox regression with forward effect selection method and stay in p-value boundary equal to 0.15. The forward method firstly computes the adjusted chi-square statistics for each variable. Then, it examines the largest computed statistics and if particular one is significant, the corresponding variable is added to the model. Once the variable is entered, it is never removed from the model. 3 out of 4 factors that appeared to be significant according to forward selection method were confirmed as the significant ones by stepwise selection method.

The results of modelling proved the possibility of prediction the survival using certain set of disease features and subjects' characteristics. Testing of global hypothesis for Beta resulted in rejecting of null hypothesis ( $\text{Beta} = 0$ ) in favor of the alternative one ( $\text{Beta} \neq 0$ ) thus it was confirmed that the models make sense and can be used to predict survival in women with breast cancer. According to obtained results, the most significant disease features and subjects characteristics appeared to be: type of multiple processes (synchronous or metachronous), presence of relapse and/or metastasis, type and combination of treatment, stage of disease.

Cancer with synchronous processes is characterized by greater aggressiveness and it reduces survival by almost 13 times compared with cancer where metachronous processes take place. Even though chemotherapy significantly increases the survival rate of patients, it also impacts the probability of relapses and metastasis occurrence, which are 16 times more likely to occur if chemotherapy was a part of treatment. This gives grounds for assumption that it has an indirect effect on survival and hence needs to be analyzed considering its negative impact on the relapses and metastasis occurrence probability, which, in turn, reduces survival by 10 times. This fact, in our opinion, introduces the need for further in-depth analysis. The significant difference between survival rates in patients with the first and third stages of cancer has been proved – the chances to survive with the disease at the first stage are almost 12 times higher than with disease at the third stage. At the same time, the difference in the survival rates in women with the disease at the second and the third stages is not so big and it is only 1.6 times. The modern method of conducting surgery compared with the standard one appeared to be capable to reduce the risk of relapses and metastases by 2.6 times, while breast conservative surgery in multiple oncological processes – by 3 times compared with mastectomy, which allows to state that both factors have a positive effect on the survival probability and reduce the risk of mortality.

Regarding subgroup models built for patients having synchronous process and patients with metachronous processes separately, an increase in the sample size is needed to assess assumed difference in factors affecting survival and to improve predictive abilities of models. This, in turn, requires additional studies during which the necessary amount of data can be collected.

**Key words:** *Cox regression, forward method, survival analysis, breast cancer, multiple primary malignant neoplasms.*

Бібліографічний опис для цитування:

Ковтун Н. В., Мотузюк І. М., Ганжа Р. О. Використання регресії Кокса для прогнозування виживаності жінок з множинними злоякісними новоутвореннями // Статистика України. 2018. № 4. С. 65–71.