УДК 519.2
DOI: https://doi.org/10.17721/1812-5409.2025/2.4

Олександр ГАЛУШКО, асп.
ORCID ID: 0009-0009-0275-6542
e-mail: health24@ukr.net
Київський національний університет імені Тараса Шевченка, Київ, Україна

ПРОГНОСТИЧНА МОДЕЛЬ ЛОГІСТИЧНОЇ РЕГРЕСІЇ ДЛЯ ОЦІНЮВАННЯ ПОКАЗАНЬ ДО ЛІКУВАЛЬНОЇ ІМПЛАНТАЦІЇ ПРИ ПРОЛАБУВАННІ ШЛУНКА У СТРАВОХІД АБО КАРДІАЛЬНУ ДІЛЯНКУ ШЛУНКА

Побудовано прогностичну модель на основі логістичної регресії для оцінювання показань до лікувальної імплантації у пацієнтів із пролабуванням шлунка у стравохід або кардіальну ділянку шлунка на основі даних 558 пацієнтів ендоскопічного відділення Кіровоградської обласної лікарні з клінічними симптомами й ендоскопічними характеристиками. Модель ураховує дев'ять предикторів: скарги на печію, регургітацію, тиск за грудиною, осиплість голосу, наявність дефектів слизової за Лос-Анджелівською класифікацією, висоту пролабування шлунка у стравохід або кардіальну ділянку шлунка, положення Z-лінії та наявність ерозій і виразок, укритих гематином. Під час побудови логістичної регресійної моделі особливу увагу приділено медичній інтерпретації отриманих коефіцієнтів, стандартних помилок, статистик Вальда, матриць варіацій, відношень шансів та іншим показникам моделі. Це дало змогу не лише оцінити статистичну значущість предикторів, а й зрозуміти їхнє клінічне значення у прогнозуванні можливості імплантації в конкретного пацієнта. Для аналізу залежностей між предикторами були використані таблиці спряженості та коефіцієнти Крамера, що допомогло виявити слабкі й помірні зв'язки між симптомами та зробити висновок про невилучення всіх предикторів з моделі. Для оцінювання пояснювальної здатності моделі використано коефіцієнти детермінації Кокса та Снелла, модифікований коефіцієнт детермінації Нейджелкера, тест Хосмера – Лемешова та класифікаційну таблицю за моделлю. Результати демонструють високу прогностичну здатність моделі (точність 98,6 %) і підтверджують клінічну значущість навіть рідкісних симптомів.

К л ю ч о в і с л о в а : ***логістична регресія; прогностична модель, медична імплантація, пролабування шлунка, ендоскопічні предиктори, клінічне рішення.***

Класифікація відповідно до AMS 2020: 62J12, 62P10.

Вступ

Задачі медичної діагностики та прогнозування належать до класу задач класифікації, де на основі множини клініко-ендоскопічних змінних необхідно визначити належність пацієнта до однієї з категорій. Математичне моделювання цього процесу вимагає використання апарату ймовірнісно-статистичних методів, що дає змогу перетворити набір емпіричних даних у формалізовану систему прийняття рішень.

Класичним підходом до таких задач є використання логістичної регресії (Hosmer, Lemeshow, & Sturdivant, 2013), що забезпечує відображення простору предикторів у простір ймовірностей за допомогою логіт-функції. Важливою властивістю логістичної регресії є можливість інтерпретації коефіцієнтів у вигляді відношення шансів, що робить її не лише інструментом прогнозування, але й засобом оцінювання значущості кожного окремого предиктора. Моделі логістичної регресії використовуються в медицині як цінний математико-статистичний апарат для виявлення хвороби Паркінсона (Verma, & Sinha, 2025); прогнозування післяопераційної смертності у хворих на рак легенів (Bernard et al., 2024); прогнозування ймовірності гострого коронарного синдрому (Hua et al., 2025) та ін. Ключові можливості застосування логістичної регресії в медицині наведені в роботі (Schober, & Vetter, 2021).

У пропонованій статті буде побудовано математичну модель, що інтегрує клінічні й ендоскопічні дані для прогнозування показань і протипоказів до імплантації діагностичних та/або терапевтичних пристроїв, зокрема систем Bravo, JSPH-1, BEST Capsule, EndoStim для стимуляції нижнього стравохідного сфінктера, капсульних датчиків Smart nanowire та ін. (Lawenko, & Lee, 2016; Li, Liu, & Tao, 2013) у пацієнтів із патологією шлунково-стравохідного з'єднання, з метою підвищення точності йоб'єктивності клінічного ухвалення рішень.

1. Постановка задачі

На основі вибірки із 558 пацієнтів ендоскопічного відділення комунального некомерційного підприємства "Кіровоградська обласна лікарня Кіровоградської обласної ради" та їхніх клінічних характеристик, лікарі, на основі детального аналізу їхнього стану, змогли оцінити можливість імплантації / не імплантації лікувального пристрою у пацієнтів із пролабуванням шлунка у стравохід або кардіальну ділянку шлунка. Основними умовами для імплантації були такі: наявність скарг на печію, відчуття регургітації та тиску за грудиною, осиплість голосу (поєднання скарг у контексті дозволу імплантації визначали лікарі); показник висоти пролабування шлунка у стравохід або кардіальну ділянку шлунка має дорівнює або бути меншим за показник Z-лінії; відсутність дефектів типу C і D за Лос-Анджелівською класифікацією оцінки рефлюкс-езофагіту; відсутність ерозій і виразок, укритих гематином у разі дефектів типу A і B й інші менш значущі характеристики, які лікарі визначали у процесі обстеження пацієнтів.

Мета пропонованої статті полягає у знаходженні моделі, яка на основі ряду отриманих клінічних показників й отриманих раніше результатів (заключень) лікарів щодо імплантації / не імплантації лікувального пристрою, зможе на основі клінічних показників нового пацієнта надати відповідь про можливість імплантації пристрою.

Проаналізуємо, наскільки логістична регресія на початкових даних здатна відтворити чіткі клінічні правила класифікації пацієнтів щодо можливості імплантації. Коефіцієнти логістичної регресії покажуть, які початкові ознаки реально впливають на результат, навіть без явного кодування правил.

2. Основні результати

У логістичній регресії предиктори (пояснювальні змінні) позначено як $x_1, x_2, \dots, x_p$, де p – кількість предикторів (у нашій задачі $p = 9$).

© Галушко Олександр, 2025

Вхідні змінні моделі (предиктори): 1) скарги на печію (бінарна змінна: 1 – наявні; 0 – відсутні) – x_1; 2) скарги на відчуття регургітації (бінарна змінна: 1 – наявні; 0 – відсутні) – x_2; 3) скарги на відчуття тиску за грудиною (бінарна змінна: 1 – наявні; 0 – відсутні) – x_3; 4) скарги на осиплість голосу (бінарна змінна: 1 – наявні; 0 – відсутні) – x_4; 5) наявність дефектів слизової, класифікованих за Лос-Анджелівською класифікацією оцінки рефлюкс-езофагіту (порядкова змінна: дефекти відсутні – 0; тип А – 1; тип В – 2; тип С – 3; тип D – 4) – x_5; 6) показник висоти пролабування шлунка в стравохід або кардіальну ділянку шлунка (кількісна змінна, см) – x_6; 7) показник розташування Z-лінії (кількісна змінна, см) – x_7; 8) наявність ерозій, укритих гематином (бінарна змінна: 1 – наявні; 0 – відсутні) – x_8; 9) наявність виразок, укритих гематином (бінарна змінна: 1 – наявні; 0 – відсутні) – x_9. Результуюча змінна Y: можливість імплантації (1 – так; 0 – ні). У нашому випадку $i = \underline{1, n}$, де $n = 558$ – кількість пацієнтів).

Нехай x_i – вектор предикторів для i-го спостереження (пацієнта): $x_i = (1, x_{i1}, x_{i2}, \dots, x_{ip})'$, де перший елемент 1 відповідає вільному члену моделі β_0. Змоделюємо ймовірність позитивного результату (успіху, настання події – імплантації пристрою) як: $p_i = f(x_i; \beta) = exp\, exp\, (x_i'\beta)\, /(1 + exp\, exp\, (x_i'\beta)\,)$, де $p_i = E[Y_i]$ – це очікувана ймовірність успіху (імплантації) для i-го спостереження (пацієнта); β – вектор кофіцієнтів моделі.

Далі логістичну функцію побудуємо як логарифмічну функцію шансів (Nguyen, 2020):

$$logit(p_i) = ln\left(\frac{p_i}{1-p_i}\right) = x_i'\beta = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip},$$

де величина $p_i/(1-p_i)$ представляє шанси на успіх (співвідношення ймовірності успіху до ймовірності невдачі).

Отже, імовірність p_i – це імовірність успіху, яка показує, наскільки ймовірно, що для i-го об'єкта (пацієнта) подія настане (буде показання до імплантації). А величина $p_i/(1-p_i)$ показує у скільки разів імовірніший успіх за невдачу (від 0 до ∞). Далі натуральний логарифм $ln\, (p_i/(1-p_i))$ перетворює шанси на шкалу від −∞ до +∞, що зручно для моделювання.

Оскільки Y_i відповідає розподілу Бернуллі з імовірністю p_i, то функція правдоподібності для n незалежних спостережень матиме вигляд

$$L(\beta) = L(p_i) = \prod_{i=1}^{n} \quad p_i^{Y_i}(1-p_i)^{1-Y_i}, \;\; p_i = \frac{exp\, (\eta_i)}{1 + exp\, (\eta_i)}, \;\; 1 - p_i = \frac{1}{1 + expexp\, (\eta_i)}, \;\; \eta_i = x_i'\beta.$$

Якщо $y_i = 1$, у добуток іде p_i. Якщо $y_i = 0$, у добуток іде $(1-p_i)$. Отже, функція $L(\beta)$ показує, "наскільки добре" модель з коефіцієнтами β описує спостережувані дані.

Зручніше працювати з логарифмом правдоподібності (лог-правдоподібністю), щоб перетворити добуток у суму:

$$l(\beta) = ln\, ln\, L(\beta) \;=\sum_{i=1}^{n} \quad [y_i\, ln\, ln\, (p_i) \;+ (1-y_i)\, ln\, ln\, (1-p_i)\,].$$

У пакеті IBM SPSS Ststistics закладені алгоритм Ньютона – Рафсона й алгоритм Фішера, що дають змогу отримати оцінки максимальної правдоподібності параметрів β. За достатньо великої вибірки розподіл $\hat{\beta}$ (оцінок коефіцієнтів) наближається до багатовимірного нормального розподілу $\hat{\beta} \sim AN(\beta, [I(\beta)]^{-1})$, де $I(\beta)$ – інформаційна матриця за Фішером, що визначає дисперсійно-коваріаційну структуру $\hat{\beta}$. Її визначають як $I(\beta) = E\left[\frac{\partial l(\beta)}{\partial \beta}\frac{\partial l(\beta)}{\partial \beta'}\right]$.

Оцінки коефіцієнтів $\hat{\beta} = l(\beta)$ перебувають в основі ітераційних чисельних методів.

Обчислюємо похідні лог-правдоподібності. Для кожного i: $\frac{\partial l}{\partial \eta_i} = y_i - p_i$.

У матричній формі градієнт записуємо як $U(\beta) = \nabla_\beta l(\beta) = X^T(y-p)$, де X – матриця $n \times p$, $p = (p_1, \dots, p_n)$, $y = (y_1, \dots, y_n)$. Матрицю Гессе розраховуємо так: $H(\beta) = \nabla_\beta^2 l(\beta) = -X^T W X$, де $W = diag(\omega_1, \dots, \omega_n)$ і $\omega_i = p_i(1-p_i)$. W – це діагональна матриця, у якої всі елементи поза головною діагоналлю дорівнюють нулю. Зауважимо, що більші ваги в контексті нашої задачі відповідатимуть пацієнтам, для яких модель менш упевнена (тобто p_i близько до 0,5). Менші ваги відповідають пацієнтам, для яких модель дуже впевнена (тобто p_i близько до 0 або 1). Іншими словами, ваги підкреслюють, наскільки "важливим" є конкретне спостереження для уточнення параметрів моделі на поточному кроці ітерації.

Формула Ньютона – Рафсона у нашому випадку матиме вигляд:

$$\beta^{(t+1)} = \beta^{(t)} - H^{-1}(\beta^{(t)})U(\beta^{(t)}) = \beta^{(t)} - \left(-X^T W^{(t)} X\right)^{-1} X^T\left(y - p^{(t)}\right) = \beta^{(t)} + \left(X^T W^{(t)} X\right)^{-1} X^T\left(y - p^{(t)}\right).$$

Уведемо робочий відгук $z^{(t)} = X\beta^{(t)} + \left[W^{(t)}\right]^{-1}\left(y - p^{(t)}\right)$. Тоді крок можна записати як $\beta^{(t+1)} = argmax_\beta\left(z^{(t)} - -X\beta\right)^T W^{(t)}\left(z^{(t)} - X\beta\right)$, що дає нормальні рівняння $\left(X^T W^{(t)} X\right)\beta^{(t+1)} = X^T W^{(t)} z^{(t)}$.

Це показує, що кожна ітерація – це крок зваженого найменших квадратів з вагами ω_i.

Асимптотична матриця варіацій $Var(\hat{\beta}) \approx I^{-1}(\hat{\beta}) \approx -H^{-1}(\hat{\beta})$. Оскільки $H(\hat{\beta}) = -X^T \widehat{W} X$, то $Var(\hat{\beta}) \approx \left(X^T \widehat{W} X\right)^{-1}$.

Стандартна помилка j-го коефіцієнту – це квадратний корінь діагонального елементу цієї матриці: $SE(\hat{\beta}_j) = \sqrt{Var(\hat{\beta}_j)}$. Стандартна помилка показує, наскільки точно оцінено вплив певного предиктора на ймовірність імплантації. Низьке значення предиктора вказує на вплив цього предиктора. Велике значення предиктора вказує, що оцінка $\hat{\beta}_j$ нестабільна, тобто даних недостатньо або вони дуже варіабельні для надійного прогнозу. У клінічній практиці навіть нестабільний предиктор може бути важливим для ухвалення рішень, особливо якщо він теоретично обґрунтований (Matsui, Le-Rademacher, & Mandrekar, 2021).

Для будь-якого предиктора x_j коефіцієнт $\hat{\beta}_j$ пов'язаний із відношенням шансів (odds ratio): $OR_j = e^{\hat{\beta}_j}$. У контексті логістичної регресії відношення показує, як зміна предиктора впливає на шанси настання події. Інтерпретація: $OR = 1$ – немає впливу предиктора; $OR > 1$ – збільшення шансу настання події за зростання предиктора; $OR < 1$ – зменшення шансу.

Наведемо критерії збіжності в ітераційному процесі в пакеті IBM SPSS Statistics: 1) $\|\beta^{(t+1)} - \beta^{(t)}\|_\infty < \varepsilon$, де $\varepsilon = 0{,}001$; 2) досягнення максимум ітерацій (maxit), де maxit = 20.

Для аналізу значущості коефіцієнтів регресії (перевірки нульової гіпотези $H_0: \beta_j = 0$) застосовуватиме статистику Вальда $W_j = \hat{\beta}_j^2 / SE(\hat{\beta}_j)^2$. Розподіл χ^2 з 1 степенем свободи дає змогу обчислити p-значення критерію. Ступені свободи для кожного коефіцієнта β_j $df = 1$, тобто тестуємо лише один параметр. Значущість (p-значення) показує ймовірність отримати таку статистику Вальда, якщо насправді $\beta_i = 0$: $p = 1 - F_{\chi^2}(W_j, df = 1)$, де F_{χ^2} – функція розподілу χ^2 з одним ступенем свободи. Мале p-значення (< 0,05) свідчить про те, що наявний предиктор x_j статистично значуще пов'язаний з імовірністю успішної імплантації. Велике p-значення показує, що недостатньо доказів, щоб стверджувати вплив предиктора, але не означає, що його можна ігнорувати клінічно.

Побудуємо логістичну регресію для перерахованих клінічних показників у статистичному пакеті IBM SPSS Statistics. У табл. 1 наведено показники рівняння логістичної регресії та їхню значущість у моделі регресії. Також подано основну інформацію щодо медичної інтерпретації показників.

Таблиця 1

Показники змінних рівняння логістичної регресії та їхня медична інтерпретація

Клінічні показники	Оцінки коефіцієнтів логістичної регресії $\hat{\beta}_j$	Середньо-квадратичні похибки, $SE(\hat{\beta}_j)$	Статистики Вальда, W_j	Ступені свободи, df	p-значення	Відношення шансів, OR_j	Медична інтерпретація показників
Скарги на печію	1,866	0,841	4,917	1	0,027	6,460	Наявність скарг збільшує шанси показань до імплантації у 6,5 раза. Значущий предиктор
Скарги на відчуття регургітації	-1,784	0,760	5,503	1	0,019	0,168	Наявність регургітації зменшує шанси у 6 разів (1/0,168 = 5,95). Змінна "перевертається" у знак "захисного фактора" штучно. Значущий предиктор
Скарги на відчуття тиску за грудиною	-17,030	19381,806	0,000	1	0,999	0,000	Коефіцієнт нестабільний. Предиктор статистично незначущий
Скарги на осиплість голосу	-0,659	1,133	0,338	1	0,561	0,517	Предиктор статистично незначущий
Наявність дефектів слизової, класифікованих за Лос-Анджелівською класифікацією оцінки рефлюкс-езофагіту	-4,432	0,737	36,188	1	0,000	0,012	Наявність дефектів різко зменшує ймовірність (майже у 83 рази). Дуже сильний і значущий предиктор
Показник висоти пролабування шлунка у стравохід або кардіальну ділянку шлунка	-1,054	0,364	8,408	1	0,004	0,348	Кожна одиниця зростання висоти пролабування шлунка у стравохід або кардіальну ділянку шлунка (см) зменшує шанси імплантації майже у 3 рази (1/0,348 = 2,87). Значущий предиктор
Показник розташування Z-лінії	0,748	0,376	3,967	1	0,046	2,114	Збільшення показника Z-лінії підвищує ймовірність імплантації більш як у 2 рази. Значущий предиктор
Наявність ерозій, укритих гематином	-23,762	8127,078	0,000	1	0,998	0,000	Технічно незначущий предиктор через рідкість явища
Наявність виразок, укритих гематином	-23,247	6186,634	0,000	1	0,997	0,000	Технічно незначущий предиктор через рідкість явища
Константа	21,066	11,324	3,461	1	0,063	$1{,}41 \times 10^9$	Велика, але не значуща

Коефіцієнт регресії $\hat{\beta}_j$ – це оцінка впливу j-ї незалежної змінної на логіт-ймовірності. Для кожної змінної x_j значення $\hat{\beta}_j$ показує, як зміна на одиницю в X_j змінює $logit(\hat{p}_i)$. Зауважимо, що для додатних значень $\hat{\beta}_j$ шанси на імплантацію зростають у $e^{\hat{\beta}_j}$-разів. У нашому випадку рівняння логістичної регресії має такий вигляд:

$$logit(\hat{p}_i) = 21{,}066 + 1{,}866x_{i1} - 1{,}784x_{i2} - 17{,}030x_{i3} - 0{,}659x_{i4} - 4{,}432x_{i5} - \\ -1{,}054x_{i6} + 0{,}748x_{i7} - 23{,}762x_{i8} - 23{,}247x_{i9}. \quad (1)$$

Проаналізуємо показники якості логістичної моделі на основі коефіцієнтів детермінації Кокса та Снелла, а також коефіцієнта Нейджелкерка. Коефіцієнт детермінації $R^2_{C\&S}$ Кокса та Снелла показує відносне поліпшення лог-правдоподібності моделі з предикторами порівняно з моделлю лише з константою (Cox, & Snell, 1989). Обраховується як

$R^2_{C\&S} = 1 - (L_0/L_m)^{\frac{2}{n}}$, де L_0 – правдободібність за нульовою моделлю (з константою); L_m – правдободібність за оціненою моделлю (з предикторами); n – кількість спостережень. Максимальне значення $R^2_{C\&S} < 1$ (часто ≈ 0,75–0,9 навіть для ідеальної моделі). Тому використовуємо модифікований коефіцієнт детермінації Нейджелкерка ($R^2_N = R^2_{C\&S}/(1 - L_0^{2/n})$), який масштабується так, щоб максимальне значення могло дорівнювати 1.

Наша модель характеризується високою пояснювальною здатністю, що підтверджується високим $R^2_N = 0{,}903$, тобто модель відтворює майже всі варіації залежної змінної.

На основі тесту Хосмера – Лемешова проаналізуємо узгодженість спостережуваних (емпіричних) й отриманих теоретичних значень моделі. Для розрахунку статистики Хосмера – Лемешова χ^2_{HL} спостережувані дані групуються у g-груп за прогнозованою ймовірністю (Hosmer, Lemeshow, & Sturdivant, 2013). Для кожної групи обчислюємо спостережувану кількість подій $\hat{E}_i$ і невідповідностей $(O_i - \hat{E}_i)$. Статистику критерію розраховуємо так: $\chi^2_{HL} = \sum_{i=1}^{g} \frac{(O_i - \hat{E}_i)^2}{\hat{E}_i\,(1 - \hat{E}_i/n_i)}$, де кількість ступенів свободи $df = g - 2$, а p-значення (значущість) розраховуємо через χ^2-розподіл. У нашому випадку на стандартному рівні 0,05 за тестом Хосмера – Лемешова модель є прийнятною ($p = 0{,}034 < 0{,}05$).

У контексті цього тесту проаналізуємо результати класифікації за отриманою моделлю. Для пацієнтів без показань до імплантації (0) модель правильно передбачила 111 зі 116 випадків, тобто 95,7 % правильних передбачень. П'ять пацієнтів були помилково передбачені як 1 (помилка I типу для цієї категорії). Для пацієнтів із показаннями до імплантації (1) модель правильно передбачила 438 із 441 випадку, тобто 99,3 % правильних передбачень. Три пацієнти були помилково класифіковані як 0 (помилка II типу). Модель правильно передбачає 98,6 % усіх випадків. Це дуже високий показник точності, що свідчить про сильну предиктивну здатність моделі.

Спробуємо проаналізувати, чи доцільно вилучати предиктори з отриманої моделі. Для цього між показниками скарг проаналізуємо залежність. Для оцінювання залежності скористаємося коефіцієнтом спряженості Крамера V (Sheskin, 2000): $V = \sqrt{\chi^2/(N \cdot (k-1))}$, де χ^2 – значення критерію Пірсона з таблиці спряженості; N – обсяг вибірки; $k = min\,(r, c)$, де r – кількість рядків, c – кількість стовпців таблиці (у нашому випадку $r = c = 2$; $V \in [0{,}1]$.

Проаналізуємо наповнення таблиць спряженості та коефіцієнти Крамера для кожної пари скарг. 1) Печія та регургітація. Регургітація майже завжди поєднується з печією (лише 14 ізольованих випадків регургітації без печії). $V = 0{,}596$. Маємо помірний статистично значущий зв'язок. Регургітація є похідним виявом рефлюксних симптомів (печії). 2) Печія та тиск за грудиною. Скарги на тиск за грудиною практично відсутні (лише 3 ізольовані випадки та 196 у поєднанні з печією). У більшості випадків (359) обидві скарги відсутні. $V = 0{,}099$. Маємо слабкий, але значущий зв'язок. У великих вибірках слабкий зв'язок часто може бути значущим (Van den Berg, 2017). 3) Печія та осиплість голосу. Скарги на осиплість голосу зустрічаються рідко (20 випадків із 558), але переважно в пацієнтів із печією (14 проти 6 випадків без печії). Це свідчить, що осиплість голосу може бути вторинним виявом гастроезофагеального рефлюксу. $V = 0{,}138$. Маємо слабкий, але значущий зв'язок. Осиплість голосу можна розглядати як ускладнений вияв печії. 4) Регургітація та тиск за грудиною. Усі випадки тиску за грудиною пов'язані з регургітацією: $V = 0{,}137$. Маємо слабкий, але значущий зв'язок. Симптоми "співіснують", проте залежність не така сильна, як у пари "печія – регургітація". 5) Регургітація й осиплість голосу. В 11 випадках регургітація поєднується з осиплістю проти 9 ізольованих випадків осиплості: $V = 0{,}151$. Маємо слабкий, але значущий зв'язок. Осиплість голосу може бути вторинним виявом регургітації. 6) Тиск за грудиною й осиплість голосу. Усі випадки осиплості виникають без тиску (20 ізольованих випадків осиплості): $V = 0{,}014$. Зв'язок статистично незначущий. Ці симптоми незалежні, можна розглядати їх окремо. Зауважимо, що між жодною парою змінних немає сильного статистичного зв'язку ($V > 0{,}6$). Отже, виводити якісь показники скарг із загальної моделі регресії не доцільно. Отже, навіть якщо зв'язки між деякими змінними слабкі, ці показники є важливими медичними предикторами і мають залишатися в моделі для збереження клінічної цінності прогнозування (Matsui, Le-Rademacher, & Mandrekar, 2021).

Що стосується предикторів наявності ерозій і виразок, укритих гематином, то у статистичному сенсі предиктори є "незначущими" через надзвичайну рідкість явища (стандартні похибки великі). Це типовий випадок проблеми "розріджених даних" (Greenland, Mansournia, & Altman, 2016). Проте їх не можна вважати "неважливими", бо вони мають високу клінічну специфічність: навіть поодинокі випадки таких уражень є показником тяжкості процесу та ризику ускладнень. У разі вилучення цих змінних модель втрачає клінічну релевантність: вона перестає "бачити" критичні, хоч і рідкісні стани, що одразу знижує її прогностичну цінність у практиці. Крім того, видалення таких змінних призводить до погіршення якості моделі. Це пов'язано з тим, що навіть рідкісні симптоми дають внесок у розмежування груп (напр., "немає патології" чи "важкі форми").

Дискусія і висновки

Отже, отримана модель логістичної регресії (1) найкраще відображає зв'язок між клінічними показниками і результуючою змінною щодо можливості імплантації лікувального пристрою. Зауважимо вплив значущих предикторів, які значно змінюють показання до імплантації пристрою. Наявність скарг на печію в пацієнта збільшує шанси показань до імплантації у 6,5 раза. Наявність регургітації зменшує шанси у шість разів. Змінна "перевертається" у знак "захисного фактора" штучно. Проте, фактично, скарги на регургітацію значно впливають на показання до імплантації. Наявність дефектів слизової за Лос-Анджелівською класифікацією різко зменшує шанси до імплантації – майже у 83 рази. Дуже сильний і значущий предиктор. Кожен сантиметр зростання висоти пролабування шлунка у стравохід або кардіальну ділянку шлунка зменшує шанси імплантації майже у три рази. Збільшення показника Z-лінії підвищує ймовірність імплантації у понад два рази.

Прогностична модель логістичної регресії є значущою за критерієм Хосмера – Лемешова. Результати демонструють високу прогностичну здатність моделі (точність 98,6 %) і підтверджують клінічну значущість навіть рідкісних симптомів.

Джерела фінансування. Фінансування дослідження здійснюється власним коштом автора та частково забезпечено Київським національним університетом імені Тараса Шевченка.

Список використаних джерел

Bernard, A., Cottenet, J., Pages, P. B., & Quantin, C. (2024). Is the Validity of Logistic Regression Models Developed with a National Hospital Database Inferior to Models Developed from Clinical Databases to Analyze Surgical Lung Cancers? *Cancers,* 16(4), 734. https://doi.org/10.3390/cancers16040734

Cox, D. R., & Snell, E. J. (1989). *Analysis of binary data* (2nd ed.). Chapman and Hall/CRC.

Greenland, S., Mansournia, M. A., & Altman, D. G. (2016). Sparse data bias: A problem hiding in plain sight. *BMJ*, 352, i1981. https://doi.org/10.1136/bmj.i1981

Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (3rd ed.). Wiley.

Hua, Y., Stead, T. S., George, A., & Ganti, L. (2025). Clinical risk prediction with logistic regression: Best practices, validation techniques, and applications in medical research. *Academic Medicine & Surgery*. https://doi.org/10.62186/001c.131964

Lawenko, R. M. A., & Lee, Y. Y. (2016). Evaluation of gastroesophageal reflux disease using the Bravo capsule pH system. *Journal of Neurogastroenterology and Motility,* 22(1), 25–30. https://doi.org/10.5056/jnm15130

Li, J. N., Liu, C. L., & Tao, X. H. (2013). Clinical utility and tolerability of JSPH-1 wireless esophageal pH monitoring system. *BMC Gastroenterology,* 13(10). https://doi.org/10.1186/1471-230X-13-10

Matsui, S., Le-Rademacher, J., & Mandrekar, S. J. (2021). Statistical models in clinical studies. *Journal of Thoracic Oncology,* 16(5), 734–739. https://doi.org/10.1016/j.jtho.2021.02.021

Nguyen, M. (2020). *A guide on data analysis.* Bookdown. https://bookdown.org/mike/data_analysis

Schober, P., & Vetter, T. R. (2021). Logistic regression in medical research. *Anesthesia & Analgesia,* 132(2), 365–366. https://doi.org/10.1213/ANE.0000000000005247

Sheskin, D. J. (2000). *Handbook of parametric and nonparametric statistical procedures* (2nd ed.). Chapman & Hall/CRC. https://dl.icdst.org/pdfs/files3/22a131fac452ed75639ed5b0680761ac.pdf

Van den Berg, R. G. (2017). *Cramér's V – What and why? SPSS Tutorials.* https://www.spss-tutorials.com/cramers-v-what-and-why/

Verma, V., & Sinha, B. (2025). Logistic regression-based detection of Parkinson's disease. *Preprints*. https://doi.org/10.20944/preprints202502.0239.v1

References

Bernard, A., Cottenet, J., Pages, P. B., & Quantin, C. (2024). Is the Validity of Logistic Regression Models Developed with a National Hospital Database Inferior to Models Developed from Clinical Databases to Analyze Surgical Lung Cancers? *Cancers,* 16(4), 734. https://doi.org/10.3390/cancers16040734

Cox, D. R., & Snell, E. J. (1989). *Analysis of binary data* (2nd ed.). Chapman and Hall/CRC.

Greenland, S., Mansournia, M. A., & Altman, D. G. (2016). Sparse data bias: A problem hiding in plain sight. *BMJ*, 352, i1981. https://doi.org/10.1136/bmj.i1981

Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (3rd ed.). Wiley.

Hua, Y., Stead, T. S., George, A., & Ganti, L. (2025). Clinical risk prediction with logistic regression: Best practices, validation techniques, and applications in medical research. *Academic Medicine & Surgery*. https://doi.org/10.62186/001c.131964

Lawenko, R. M. A., & Lee, Y. Y. (2016). Evaluation of gastroesophageal reflux disease using the Bravo capsule pH system. *Journal of Neurogastroenterology and Motility,* 22(1), 25–30. https://doi.org/10.5056/jnm15130

Li, J. N., Liu, C. L., & Tao, X. H. (2013). Clinical utility and tolerability of JSPH-1 wireless esophageal pH monitoring system. *BMC Gastroenterology,* 13(10). https://doi.org/10.1186/1471-230X-13-10

Matsui, S., Le-Rademacher, J., & Mandrekar, S. J. (2021). Statistical models in clinical studies. *Journal of Thoracic Oncology,* 16(5), 734–739. https://doi.org/10.1016/j.jtho.2021.02.021

Nguyen, M. (2020). *A guide on data analysis.* Bookdown. https://bookdown.org/mike/data_analysis/

Schober, P., & Vetter, T. R. (2021). Logistic regression in medical research. *Anesthesia & Analgesia,* 132(2), 365–366. https://doi.org/10.1213/ANE.0000000000005247

Sheskin, D. J. (2000). *Handbook of parametric and nonparametric statistical procedures* (2nd ed.). Chapman & Hall/CRC. https://dl.icdst.org/pdfs/files3/22a131fac452ed75639ed5b0680761ac.pdf

Van den Berg, R. G. (2017). *Cramér's V – What and why? SPSS Tutorials.* https://www.spss-tutorials.com/cramers-v-what-and-why/

Verma, V., & Sinha, B. (2025). Logistic regression-based detection of Parkinson's disease. *Preprints*. https://doi.org/10.20944/preprints202502.0239.v1

Отримано редакцією журналу / Received: 05.08.25
Прорецензовано / Revised: 03.10.25
Схвалено до друку / Accepted: 10.10.25

Oleksandr HALUSHKO, PhD Student
ORCID ID: 0009-0009-0275-6542
e-mail: health24@ukr.net
Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

PROGNOSTIC LOGISTIC REGRESSION MODEL FOR ASSESSING INDICATIONS FOR THERAPEUTIC IMPLANTATION WITH GASTRIC PROLAPSE INTO THE ESOPHAGUS OR THE CARDIAC PART OF THE STOMACH

In this study, a prognostic model based on logistic regression was developed to assess the indications for therapeutic implantation in patients with gastric prolapse into the esophagus or the gastric cardia. The analysis was performed on clinical and endoscopic data from 558 patients of the Endoscopy Department of the Kirovohrad Regional Hospital. The model incorporated nine predictors: complaints of heartburn, regurgitation, retrosternal pressure, hoarseness, presence of mucosal defects according to the Los Angeles classification, height of gastric prolapse into the esophagus or the gastric cardia, the position of the Z-line, and the presence of hematin-covered erosions or ulcers. Special attention was devoted to the medical interpretation of the estimated coefficients, standard errors, Wald statistics, variance–covariance matrices, odds ratios, and other model diagnostics. This approach enabled not only the evaluation of the statistical significance of the predictors but also the understanding of their clinical relevance in predicting the feasibility of implantation in individual patients. To analyze dependencies among predictors, contingency tables and Cramér's V coefficients were applied, which revealed weak to moderate associations between symptoms and justified retaining all predictors in the model. To assess the significance of the model, the Cox and Snell coefficients of determination, the modified Nagelker coefficient of determination, the Hosmer – Lemeshov test, and the classification table for the model were used. The results demonstrate a high predictive ability of the model (accuracy of 98.6%) and confirm the clinical significance of even rare symptoms.

Keywords: ***logistic regression, prognostic model, medical implantation, gastric prolapse, endoscopic predictors, clinical decision.***

Автор заявляє про відсутність конфлікту інтересів. Спонсори не брали участі в розробленні дослідження; у зборі, аналізі чи інтерпретації даних; у написанні рукопису; в рішенні про публікацію результатів.

The author declares no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses or interpretation of data; in the writing of the manuscript; in the decision to publish the results.