

DOI: 10.31793/1680-1466.2026.31-1.97

Ідентифікація випадків злоякісних новоутворень серед учасників епідеміологічних досліджень віддалених наслідків аварії на Чорнобильській АЕС

Є.Л. Горох¹,
В.М. Шпак²,
О.В. Сумкіна¹,
А.Ю. Рижов³,
З.П. Федоренко¹,
Л.О. Гулак¹,
О.В. Лапікура²,
М.Д. Тронько²

¹ДНП «Національний інститут раку»

²ДУ «Інститут ендокринології та обміну речовин ім. В.П. Комісаренка НАМН України»

³Київський національний університет імені Тараса Шевченка

Резюме. Проспективні когортні дослідження, у тому числі дослідження віддалених наслідків аварії на Чорнобильській атомній електростанції (ЧАЕС), потребують тривалого спостереження за учасниками з метою реєстрації їх життєвого стану та медичних подій, зокрема захворювань на злоякісні новоутворення (ЗН). Одним із використаних методів є автоматизоване зіставлення даних (record linkage) когорти Українсько-Американської програми «Науковий проєкт вивчення раку та інших захворювань щитоподібної залози в Україні в результаті аварії на Чорнобильській АЕС» (УкрАм когорти) та медичних записів Національного канцер-реєстру України (НКРУ). **Мета.** Описати методи ідентифікації в базі даних НКРУ захворювань на ЗН серед членів радіологічно значущих когорт, досвід застосування record linkage в провідних пост-чорнобильських дослідженнях та деякі наукові результати, отримані з використанням подібних технологій. **Матеріал і методи.** Використані дані НКРУ за період з 1998 по 2024 рік. Автоматизоване зіставлення даних проводилося з використанням алгоритмів та програмного забезпечення, розроблених НКРУ. **Результати.** Представлений опис методології, процесів та результатів автоматизованого зіставлення даних (record linkage) для пошуку випадків захворювання на ЗН щитоподібної залози (ЩЗ) та інших локалізацій серед членів УкрАм когорти та учасників інших епідеміологічних досліджень. **Висновки.** Використання електронних баз даних НКРУ та автоматизованих методів record linkage дозволило ефективно виявляти захворювання на ЗН серед членів УкрАм когорти протягом усього періоду дослідження. Отримана інформація дає змогу кількісно оцінити вплив віддалених наслідків аварії на ЧАЕС на онкологічну захворюваність досліджуваної когорти.

Ключові слова: Національний канцер-реєстр України, record linkage, когортні дослідження, рак щитоподібної залози.

Когортні дослідження, зокрема аналіз впливу віддалених наслідків аварії на ЧАЕС, традиційно стикаються з низкою проблем, пов'язаних із необхідністю тривалого спостереження за учасниками дослідження. Збір актуальних відомостей щодо членів Когорти є надзвичайно трудомістким процесом, особливо це стосується медичних даних. Пацієнти під час анкетування не завжди можуть надати вірогідну та документально підтверджену інформацію щодо виявлених станів та деталей діагнозу, зокрема онкологічного. Відповідна первинна медична документація є фрагментованою, її пошук та отримання доступу до неї є складним завданням.

Метою цієї роботи є охарактеризувати проблему збору даних про онкопатологію при реалізації довготривалих радіоепідеміологічних досліджень, методи ідентифікації осіб із діагнозом ЗН у базі даних НКРУ для членів визначених популяційних груп (епідеміологічних когорт) та вказати деякі наукові результати, отримані із застосуванням процедур автоматизованого зіставлення даних (record linkage).

Матеріал і методи

Централізованим джерелом даних щодо онкологічної захворюваності в Україні є НКРУ. База даних НКРУ почала формуватися в 1990-х роках на основі загальнодержавної системи реєстрації онкологічної інформації, запровадженої в СРСР в 1950-х роках. Із затвердженням Наказу Міністерства охорони здоров'я України № 10 від 22.01.1996 року «Про створення національного канцер-реєстру України», НКРУ набув офіційного статусу. Система передбачає реєстрацію та агрегацію інформації за регіональним принципом; цю функцію виконують регіональні канцер-реєстри, які працюють як структурні підрозділи онкологічних центрів. Виконує роль центральної ланки та формує централізовану загальнодержавну базу НКРУ, а також здійснює організаційно-методичний та технічний супровід регіональних канцер-реєстрів відповідний підрозділ ДНП «Національний інститут раку». Організаційна структура та принципи діяльності з реєстрації раку в Україні наразі визначені наказом Міністерства охорони здоров'я України від 01.10.2013 № 845 «Про систему онкологічної допомоги населенню України».

Більшість регіонів України до 1998 року були охоплені автоматизованою технологією НКРУ. Зокрема, серед регіонів, найбільш постраждалих від наслідків аварії на ЧАЕС, повна база даних

щодо випадків захворювання на ЗН була сформована в Житомирській області – починаючи з 1986 р., у Чернігівській обл. – із 1993 р., у Київській обл. та м. Києві – із 1997 р. Починаючи з 2002 р., база даних НКРУ охоплює випадки захворювання на ЗН на всій території України. Станом на початок 2025 року, база даних НКРУ містить інформацію про більш ніж 5 мільйонів випадків захворювання на ЗН. Кількість зареєстрованих у НКРУ ЗН серед осіб, що зазнали радіаційного опромінення внаслідок аварії на ЧАЕС, наведена в **таблиці** в розподілі за віковими групами на момент аварії на ЧАЕС.

У НКРУ реалізовані правила щодо реєстрації онкологічних діагнозів, запроваджені Міжнародною агенцією з дослідження раку (International Agency on Research of Cancer, IARC), Міжнародною асоціацією канцер-реєстрів International Association of Cancer Registries, IACR), Європейською мережею канцер-реєстрів (European Network of Cancer Registries, ENCR), членом якої НКРУ є з 1998 року [1-3]. Деперсоналізовані дані НКРУ надавалися для підготовки публікацій Всесвітньої організації охорони здоров'я «Cancer Incidence in Five Continents» (vol. X, XI, XII) та «International Incidence of Childhood Cancer» (vol. 3) [4-7], а також залучені в кількох міжнародних проєктах [8-11]. Проведене у 2018 році дослідження щодо повноти та якості даних НКРУ підтвердило, що НКРУ залишається надійним джерелом онкологічної інформації в Україні та відповідає рівню сучасних міжнародних вимог до канцер-реєстрів [12].

Для дедублікації власних даних, а також для співставлення даних НКРУ з іншими джерелами, у НКРУ були розроблені автоматизовані технології зіставлення даних, які містять ідентифікаційну інформацію (засоби record linkage). Ці засоби дозволяють виявляти записи різних баз даних, які належать одній особі, за принципом «перетину множин». Підхід базується на автоматизованому зіставленні однакових за змістом реквізитів, які ідентифікують пацієнта в різних наборах даних, та обчисленні «ваги зв'язку» для кожної пари записів, яка умовно відповідає ймовірності того, що ці записи належать одній особі. Математичну теорію цього процесу було запропоновано ще в 1960-х роках [13, 14]. Із розвитком технічних можливостей та поширенням баз даних такий підхід почав широко застосовуватися в різних сферах, включаючи епідеміологічні дослідження, які потребують інтеграції да-

Таблиця. Кількість записів у базі даних НКРУ на початок 2025 року

Table. Number of records in the database of the National Cancer Registry of Ukraine at the beginning of 2025

Група населення Population group	Роки взяття на облік Years of registration		Кількість записів Number of records
	1990-2021	2022-2024	
ЗН усіх локалізацій, C00-C96 All malignant tumours, C00-C96			
Україна Ukraine	4 946 311	361 843	5 308 154
Опромінені в дорослому віці (народжені в 1967 році чи раніше) Exposed in adulthood (born in 1967 or earlier)	4 611 952	278 421	4 890 373
Опромінені в дитячому чи підлітковому віці (народжені в 1968-1986 роках) Exposed in childhood or adolescence (born in 1968-1986)	280 797	71 031	351 828
Народжені після аварії на ЧАЕС (народжені в 1987 році чи пізніше) Born after the Chernobyl accident (born in 1987 or later)	53 562	12 391	65 953
ЗН ЩЗ, C73 Thyroid malignant tumours, C73			
Україна Ukraine	84 795	9340	94 135
Опромінені в дорослому віці (народжені в 1967 році чи раніше) Exposed in adulthood (born in 1967 or earlier)	60 977	3212	64 189
Опромінені в дитячому чи підлітковому віці (народжені в 1968-1986 роках) Exposed in childhood or adolescence (born in 1968-1986)	19 801	4253	24 054
Народжені після аварії на ЧАЕС (народжені в 1987 році чи пізніше) Born after the Chernobyl accident (born in 1987 or later)	4017	1875	5892

них із різних джерел [15-17].

Основними етапами цього процесу, запровадженого в НКРУ, є:

1. Підготовка та стандартизація даних.

Формувалися вибірки даних НКРУ та інших джерел (УкрАм когорта), які відповідають предмету дослідження та містять реквізити, які дозволяють ідентифікувати пацієнта. Зокрема, при зіставленні даних НКРУ та УкрАм когорти використовувалися наступні реквізити, які є спільними для обох джерел: прізвище, ім'я, по-батькові, дата народження, місце проживання (код регіону), адреса (у вигляді текстової строки).

Стандартизація даних передбачає приведення всіх реквізитів до уніфікованої форми, яка допускає подальшу автоматизовану обробку. Зокрема, на цьому етапі використовувалися таблиці автоматизованого російсько-українського перекладу імен, по батькові та частково прізвищ, оскільки значна частина ретроспективних даних у деяких регіонах у 1990-х та на початку 2000-х років вносилися російською мовою. Також стандартизації підлягали скорочені імена, варіації написання імен та інше з використанням та-

блиць, побудованих на основі даних НКРУ.

2. Автоматизований пошук відповідних записів.

Для оптимізації автоматизованої обробки формувалися «пакетні вирази» – умови, які дозволяють шукати відповідності для кожного запису лише в межах відповідного «пакету» іншої бази даних. Як пакети були використані: рік народження, комбінація перших трьох букв прізвища + перша буква імені.

У разі помилки в одному з ключових реквізитів пакета наявність більш ніж одного пакета дозволяла знайти відповідність при автоматизованому опрацюванні записів іншого пакету.

Для кожного запису УкрАм когорти в процесі обробки формувалися пакети відповідних записів вибірки НКРУ та для кожного з записів кожного з пакетів обчислювалася умовна вага відповідності записів – «вага зв'язку» за допомогою спеціальних функцій порівняння, розроблених для кожного з реквізитів з урахуванням специфіки вихідних даних, зокрема:

- для ПІБ – функції нечіткого порівняння текстових рядків (адаптовані для української мови варіації Levenstein distance), що дозво-

ляє знаходити відповідності в записах з похибками;

- для адрес – комбінація попарного аналізу лексем, з яких складається адреса, та підходу «Level 2 string comparing» дозволила обробляти неструктуровані строкові адреси та отримувати ненульову вагу порівняння для численних випадків перейменування вулиць та населених пунктів за рахунок інших складників адреси;
- для регіонів – функція, яка враховує зміну району в межах однієї області, як частковий збіг, оскільки такі міграції є найбільш поширеними;
- для дат народження – комбінована функція, яка дозволяє врахувати можливість помилки в даті та/або неповну дату народження (внесений тільки рік), що є характерним для ретроспективних «брудних» даних.

Сумарна вага зв'язку формувалася як сума результатів функцій порівняння окремих реквізитів, з урахуванням ймовірностей випадкового співпадіння реквізитів у різних осіб та неспівпадіння (зміни або похибки) реквізитів однієї особи в різних базах даних. Так, при тривалому спостереженні та обробці рознесених у часі записів зміни адрес або прізвищ жінок через заміжжя є доволі ймовірною подією.

Для обробки даних УкрАм когорти, поновлюваних у результаті проведення додаткових циклів анкетування, та урахування можливих змін адрес та прізвищ, був використаний підхід із послідовним відпрацюванням автоматизованого алгоритму на кожній із комбінацій даних УкрАм когорти (загалом оброблялося до 19 варіацій), та з використанням максимальної отриманої ваги зв'язку в якості фінальної.

Результат цього етапу – впорядкований за вагою зв'язку перелік пар записів НКРУ та УкрАм когорти, які можливо належать одній особі. У разі, якщо сумарна вага зв'язку пари записів виявлялася нижчою за «нижню границю», що фактично виключало можливість ідентифікації цих записів як таких, що належать одній особі, таку пару не включали до результатів та подальшої обробки. Пари записів із максимальною вагою, яку давало фактично повне співпадіння всіх ключових реквізитів, вважалися такими, що належать одній особі. Враховуючи обмежену кількість порівнюваних ідентифікаційних реквізитів, «брудні» та неповні дані обох джерел і тривалий етап спостереження, більшість співпадінь, виявлених на етапі автоматизованої обробки, були віднесені

до «сірої зони» (можливі зв'язки) та потребували додаткового експертного опрацювання.

3. Опрацювання результатів.

Усі пари записів, визначені як «можливі зв'язки», переглядалися з метою додаткової ідентифікації, з урахуванням можливих додаткових факторів, які не могли бути оброблені автоматизовано та призвели до суттєвого зниження ваги зв'язку, як-от: численні похибки; зменшувальні імена, скорочення або варіації імен, не враховані в таблицях відповідності на етапі стандартизації даних; можливі перейменування вулиць та ін. Пари, визнані на цьому етапі як «ймовірні зв'язки», для гарантованої ідентифікації яких недостатньо відомостей, що містяться в базах даних, передавалися для подальшого уточнення та ідентифікації локальним персоналом або шляхом додаткового контакту з учасником дослідження, або через звернення до закладу охорони здоров'я, де за даними НКРУ пацієнт отримував лікування. Таким чином, комбінація автоматизованого пошуку відповідностей та експертного опрацювання результатів дозволяє мінімізувати ризики втрати випадків захворювань серед членів УкрАм когорти та виключити можливість хибної ідентифікації.

Дослідження виконані в рамках Українсько-Американської програми «Науковий проєкт вивчення раку та інших захворювань щитоподібної залози в Україні в результаті аварії на Чорнобильській АЕС».

Результати та обговорення

Використання автоматизованих баз даних та технологій, їх зіставлення дозволило ідентифікувати випадки захворювання на ЗН серед членів УкрАм когорти та інших досліджуваних когорт і отримати з бази даних НКРУ додаткову медичну інформацію щодо них. Ці технології довели свою ефективність в умовах «брудних даних» та обмежених ресурсів, характерних для епідеміологічних досліджень в Україні.

Результати record linkage з базою даних НКРУ дали змогу отримати низку важливих наукових результатів стосовно захворюваності та радіаційного ризику онкологічної патології серед різних досліджуваних груп населення, що зазнало радіаційного опромінення внаслідок аварії на ЧАЕС.

Так SIR аналіз випадків нетиреоїдної онкопатології серед 13 203 членів УкрАм когорти, опромінених у віці 0-18 років, не виявив статистично значущої надлишкової захворюваності в період до 23 років після опромінення [18]. Беручи до уваги те, що мінімальний латентний період для солідних

радіогенних карцином, крім раку ЩЗ, оцінюють у 10-15 років, необхідними є подальші тривалі спостереження над вказаною когортою для отримання більш обґрунтованих висновків щодо віддалених наслідків радіаційного опромінення.

За результатами пошуку випадків захворювання на лейкемію (Lymphoid leukemia) (ICD10: C91) серед когорти ліквідаторів, аналіз 137 випадків показав лінійну залежність відносного надлишкового ризику від дози при $ERR/Gy = 1,26$ (95% ДІ: 0,03-3,58) [19].

Пошук випадків раку ЩЗ (ICD10: C73) серед когорти ліквідаторів виявив у базі даних НКРУ 149 захворювань, діагностованих у період з 1988 до 2012 року. Це дало змогу провести дослідження випадок-контроль та оцінити надлишковий відносний ризик для раку ЩЗ $EOR/Gy=0,40$ (95% ДІ: 0,05-1,48; p -value = 0,118) [20].

26 випадків раку молочної залози (ICD10: C50), виявлені методами record linkage серед 2631 жінки, що годували груддю чи були вагітні в травні-червні 1986 року стали основою SIR аналізу захворюваності відповідного населення на рак молочної залози в період із 1998 до 2016 року. Аналіз не виявив надлишкової захворюваності в жінок, що були вагітними, $SIR=0,75$ (95% ДІ: 0,44-1,18), проте вказав на потенційну наявність надлишкового ризику в жінок, що годували груддю, $SIR=1,30$ (95% ДІ: 0,40-3,01) [21].

Дві УкрАм когорти, що включають осіб, опроміненіх у дитячому чи підлітковому віці – основна когорта, близько 13 200 осіб – та осіб, опроміненіх пренатально (в утробі матері, *in utero*) – когорта *in utero*, близько 2580 осіб – відносять до радіологічно значущих когорт. Когорти спостерігаються понад 25 років та понад 20 років відповідно, члени обох когорт мають надійні індивідуальні оцінки доз опромінення [22]. У 2023 році ДУ «Інститут ендокринології та обміну речовин ім. В.П. Комісаренка НАМН України» та Національний інститут раку, США (National Cancer Institute, USA) підписали угоду, що передбачає пасивний скринінг двох когорт до 2033 року. Важливим елементом пасивного скринінгу є періодичний record linkage з базою даних НКРУ, що дозволить аналізувати захворюваність на онкопатологію та оцінити радіаційний ризик у віддалений період після опромінення (40 та більше років після аварії) [22].

У період створення та розвитку численних державних, клінічних реєстрів та інших електронних баз даних як медичного, так і адміні-

стративного спрямування record linkage залишається актуальним інструментом для реалізації масштабних епідеміологічних досліджень.

Програмні інструменти record linkage, створені НКРУ, можуть бути адаптовані для пошуку інформації в інших реєстрах, наприклад клінічний реєстр ДУ «Інститут ендокринології та обміну речовин ім. В.П. Комісаренка НАМН України» чи державний реєстр померлих осіб, що може суттєво доповнити епідеміологічні дослідження медичних наслідків аварії на ЧАЕС.

З іншого боку багаторічний досвід успішного застосування record linkage з базою даних НКРУ в цілому спектрі постчорнобильських досліджень може бути корисним при плануванні та реалізації інших довготривалих епідеміологічних досліджень, наприклад щодо працівників ядерної промисловості, населення, що проживає навколо діючих ядерних реакторів, певних контингентів військовослужбовців.

Висновки

1. Використання електронних баз даних Національного канцер-реєстру України та автоматизованих методів record linkage дозволило ефективно виявляти випадки злоякісних новоутворень серед членів Українсько-Американської тиреоїдної когорти, когорти ліквідаторів, та в низці інших епідеміологічних досліджень.

2. Ідентифікована інформація дозволила отримати низку принципових наукових результатів щодо оцінок радіаційного ризику лейкемії, раку щитоподібної залози, раку молочної залози та іншої онкологічної патології для населення України, що зазнало опромінення внаслідок аварії на Чорнобильській АЕС.

3. Довготривале спостереження чи спостереження протягом усього життя членів радіологічно значущих когорт (когорти ліквідаторів, Українсько-Американська тиреоїдна когорта) як важливий елемент дослідження включає «пасивний» скринінг онкологічної патології, що передбачає періодичний record linkage демографічних даних членів когорти з базою даних Національного канцер-реєстру України.

4. Розроблені програмні інструменти record linkage можуть бути адаптовані для пошуку інформації про членів когорти в інших реєстрах, наприклад клінічних реєстрах чи базі даних померлих.

5. Багаторічний досвід успішного застосування record linkage з базою даних Національного канцер-реєстру України в постчорнобильських

дослідженнях може бути узагальненим та ко-рисним при плануванні та реалізації інших дов-готривалих епідеміологічних досліджень.

Перспективи подальших досліджень

У віддалений період після аварії (40 та більше років) застосування методів record linkage для членів УкрАм когорти та інших досліджуваних груп (жінки, що годували груддю під час аварії, матері членів когорти, тощо) є ефективним ін-струментом для моніторингу захворюваності на ЗН та подальшого аналізу радіаційних ризиків.

Список використаної літератури

- World Health Organization. International classification of diseases for oncology (ICD-O) – 3rd edition, 1st revision. Geneva, Switzerland: WHO Press; 2013. Available from: <https://iris.who.int/server/api/core/bitstreams/f63d7652-1254-47f4-8ab7-3c65ad60d57d/content> [Accessed 12th December 2025]
- Tyczynski JE, D emaret E, Parkin D. (eds). Standards and guidelines for cancer registration in Europe. The ENCR Recommendations. Vol. 1. IARC Technical publication No. 40. Lyon: IARC; 2003. Available from: <https://publications.iarc.who.int/Book-And-Report-Series/Iarc-Technical-Publications/Standards-And-Guidelines-For-Cancer-Registration-In-Europe-2003> [Accessed 12th December 2025].
- The European Network of Cancer Registries Recommendations. Available from: <https://www.enrc.eu/ENCR-Recommendations> [Accessed 10th December 2025].
- Forman D, Bray F, Brewster DH, Gombe Mbalawa C, Kohler B, Pi neros M, Steliarova Foucher E, Swaminathan R, Ferlay J, (eds). Cancer incidence in five continents, Vol. X. IARC scientific publication No. 164. Lyon: IARC; 2014. Available from: <https://publications.iarc.who.int/Book-And-Report-Series/Iarc-Scientific-Publications/Cancer-Incidence-In-Five-Continents-Volume-X-2014> [Accessed 12th December 2025].
- Bray F, Colombet M, Mery L, Pi neros M, Znaor A, Zanetti R, Ferlay J, editors (2021). Cancer incidence in five continents, Vol. XI. IARC Scientific publication No. 166. Lyon: International Agency for Research on Cancer. Available from: <https://publications.iarc.who.int/Book-And-Report-Series/Iarc-Scientific-Publications/Cancer-Incidence-In-Five-Continents%20Volume-XI-2021> [Accessed 15th December 2025].
- Bray F, Colombet M, Aitken JF, Bardot A, Eser S, Galceran J, et al., editors (2024). Cancer incidence in five continents, Vol. XII. IARC scientific publication No. 169. Lyon, France: International Agency for Research on Cancer. Available from: <https://publications.iarc.who.int/641> [Accessed 15th December 2025].
- Steliarova-Foucher E, Colombet M, Ries LAG, Moreno F, Dolya A, Shin HY, Hesselting P, Stillier CA, editors (2025). International incidence of childhood cancer, volume III (IARC Scientific Publication No. 170). Lyon, France: International Agency for Research on Cancer. Available from: <https://publications.iarc.who.int/658> [Accessed 15th December 2025].
- European Commission. European Cancer Information System (2002-2012). Available from: <https://ecis.jrc.ec.europa.eu/> [Accessed 3rd December 2025].
- Global cancer observatory. Available from: <https://gco.iarc.fr/today/home> [Accessed 5th December 2025].
- American Cancer Society. Global cancer facts & figures. Available from: <https://www.cancer.org/research/cancer-facts-statistics/global.html> [Accessed 8th December 2025].
- Global Health Data Exchange. Global burden of disease study (2000-2012). Available from: <http://ghdx.healthdata.org/geography/ukraine> [Accessed 4th December 2025].
- Ryzhov A, Bray F, Ferlay J, Fedorenko Z, Goulak L, Gorokh Y, et al. Evaluation of data quality at the National Cancer Registry of Ukraine. *Cancer Epidemiol.* 2018 Apr;53:156-65. doi: 10.1016/j.canep.2018.02.002.
- Newcombe HB. Record linking: the design of efficient systems for linking records into individual and family histories. *Am J Hum Genet.* 1967 May;19(3 Pt 1):335-59.
- Fellegi IP, Sunter AB. A theory of record linkage. *J Am Statist*

- Assoc. 1969;64:1183-1210.
- Howe GR, Lindsay J. A generalized iterative record linkage computer system for use in medical follow-up studies. *Comput Biomed Res.* 1981 Aug;14(4):327-40. doi: 10.1016/0010-4809(81)90004-5.
- Pereira TF, Aranha VJ, Waldvogel BC, da Costa AM, Tavares Guerreiro Fregnani JH. Deterministic linkage for improving follow-up time in a Brazilian population-based cancer registry. *Sci Rep.* 2023 Mar 24;13(1):4816. doi: 10.1038/s41598-023-31303-6.
- Hejblum BP, Weber GM, Liao KP, Palmer NP, Churchill S, Shadick NA, et al. Probabilistic record linkage of de-identified research datasets with discrepancies using diagnosis codes. *Sci Data.* 2019 Jan 8;6:180298. doi: 10.1038/sdata.2018.298.
- Hatch M, Ostroumova E, Brenner A, Federenko Z, Gorokh Y, Zvinchuk O, et al. Non-thyroid cancer in Northern Ukraine in the post-Chernobyl period: short report. *Cancer Epidemiol.* 2015 Jun;39(3):279-83. doi: 10.1016/j.canep.2015.02.002.
- Zablotska LB, Bazyka D, Lubin JH, Gudzenko N, Little MP, Hatch M, et al. Radiation and the risk of chronic lymphocytic and other leukemias among Chernobyl cleanup workers. *Environ Health Perspect.* 2013 Jan;121(1):59-65. doi: 10.1289/ehp.1204996.
- Gudzenko N, Mabuchi K, Brenner AV, Little MP, Hatch M, Drozdovitch V, et al. Risk of thyroid cancer in Ukrainian cleanup workers following the Chernobyl accident. *Eur J Epidemiol.* 2022 Jan;37(1):67-77. doi: 10.1007/s10654-021-00822-9.
- Vij V, Shpak N, Zamotayeva G, Lapikura O, Ryzhov A, Gorokh E, et al. Breast cancer risk in Ukrainian women exposed to Chernobyl fallout while pregnant or lactating: standardized incidence ratio analysis, 1998 to 2016. *Eur J Epidemiol.* 2022 Nov;37(11):1195-1200. doi: 10.1007/s10654-022-00913-1.
- Тронько МД, Замотаєва ГА, Шпак ВМ, Терехова ГМ, Ляпікура ОВ, Діденко ЮА. Хронологія виконання та перспективи Українсько-Американського тиреоїдного Проекту. *Ендокринологія.* 2025 Березень 30;30(1):78-88 (Tronko MD, Zamotayeva HA, Shpak VM, Terekhova HM, Lapikura OV, Didenko YuA. Chronology of implementation and prospects of the Ukrainian-American Thyroid Project. *Endokrynologia.* 2025 Mar 30;30(1):78-88. Ukrainian). doi: 10.31793/1680-1466.2025.30-1.78.

Список скорочень

ЗН – злоякісне новоутворення

НКРУ – Національний канцер-реєстр України

УкрАм когорта – когорта Українсько-Американської програми «Науковий проект вивчення раку та інших захворювань щитоподібної залози в Україні в результаті аварії на Чорнобильській АЕС»

ЧАЕС – Чорнобильська атомна електростанція

ЩЗ – щитоподібна залоза

Identification of cases of malignant tumors among participants of epidemiology studies of the long-term consequences of the Chonobyl accident

E.L. Gorokh¹, V.M. Shpak², O.V. Sumkina¹, A.Yu. Ryzhov³, Z.P. Fedorenko¹, L.O. Gulak², O.V. Lapikura², M.D. Tronko²

¹Nonprofit Organization «National Cancer Institute»

²State Institution «V.P. Komisarenko Institute of Endocrinology and Metabolism of the National Academy of Medical Sciences of Ukraine»

³Taras Shevchenko National University of Kyiv

Abstract. Prospective cohort studies, including those of the long-term consequences of the Chernobyl nuclear power plant (ChNPP) accident, require long-term follow-up of participants to record their vital status and medical events, including malignant neoplasms (MN). One of the methods used is automated data linkage (record linkage) of the cohort of the Ukrainian-American program «Scientific project to study cancer and other thyroid diseases in Ukraine as a result of the Chernobyl accident» (UkrAm cohort) and medical records of the National Cancer Registry of Ukraine (NCRU). Objective. To describe methods for identifying in the Na-

tional Cancer Register of Ukraine database malignant neoplasms among members of radiologically significant cohorts, experience in applying record linkage in leading post-Chornobyl studies, and some scientific results obtained using similar technologies. **Material and methods.** Records of the NCRU database for the period 1998-2024 were used. Automated data matching was performed using algorithms and software developed in the NCRU. **Results.** A description of the methodology, processes, and results of automated data matching (record linkage) for searching cases of malignant neoplasm of thyroid or other sites among members of the UkrAm cohort and participants of other epidemiological studies is presented. **Conclusions.** The use of electronic databases of the National Cancer Registry of Ukraine and automated record linkage technique allowed the effective detection of malignant neoplasms among members of the UkrAm cohort throughout the entire study period. The information obtained makes it possible to quantitatively evaluate the impact of the long-term consequences of the Chornobyl accident on the oncological morbidity of the studied cohort.

Keywords: National Cancer Registry of Ukraine, record linkage, cohort studies, thyroid cancer.

Для цитування: Горох ЄЛ, Шпак ВМ, Сумкіна ОВ, Рижов АЮ, Федоренко ЗП, Гулак ЛО, Лапікура ОВ, Тронько МД. Ідентифікація випадків злоякісних новоутворень серед учасників епідеміологічних досліджень віддалених наслідків аварії на Чорнобильській АЕС. *Ендокринологія.* 2026;31(1):97-103. DOI: 10.31793/1680-1466.2026.31-1.97.

Адреса для листування: Шпак Віктор Михайлович, v.m.shpak@gmail.net; ДУ «Інститут ендокринології та обміну речовин ім. В.П. Комісаренка НАМН України», вул. Вишгородська, 69, м. Київ, 04114, Україна.

Відомості про авторів: Горох Євгеній Леонідович, канд. техн. наук, старший науковий співробітник ДНП «Національний інститут раку», ORCID: 0009-0000-9063-5824; Шпак Віктор Михайлович, старший науковий співробітник відділу з питань медичних наслідків аварії на ЧАЕС та міжнародних відносин ДУ «Інститут ендокринології та обміну речовин ім. В.П. Комісаренка НАМН України», ORCID: 0000-0002-6983-5490; Сумкіна Олена Володимирівна, провідна інженерка ДНП «Національний інститут раку», ORCID: 0000-0003-2052-9043; Рижов Антон Юрійович, канд. фіз.-мат наук, доцент Київського національного університету імені Тараса Шевченка, ORCID: 0000-0003-4099-0742; Федоренко Зоя Павлівна, канд. мед. наук, лікарка-статистик ДНП Національний інститут раку», ORCID: 0000-0003-2322-9695; Гулак Людмила Олегівна, канд. техн. наук, провідна інженерка відділу з питань медичних наслідків аварії на ЧАЕС та міжнародних відносин ДНП «Національний інститут раку», ORCID: 0009-0009-2086-0335; Лапікура Олесь Валерійович, співробітник Центру координації даних ДУ «Інститут ендокринології та обміну речовин ім. В.П. Комісаренка НАМН України», ORCID: 0000-0001-7629-2933; Тронько Микола Дмитрович, д-р мед. наук, проф., чл.-кор. НАН України, акад. НАМН України, завідувач відділу фундаментальної та прикладної ендокринології, в.о. директора ДУ «Інститут ендокринології та обміну речовин ім. В.П. Комісаренка НАМН України», ORCID: 0000-0001-7421-0981.

Особистий внесок: Горох ЄЛ. – ідея, загальне керівництво, розробка інструментів record linkage, дизайн дослідження, аналіз результатів, написання, оформлення, підготовка до друку статті; Шпак В.М. – участь у розробці концепції статті, участь у написанні рукопису; Сумкіна О.В. – формування вибірок даних НКРУ, участь у написанні рукопису; Рижов А.Ю. – участь у розробці інструментів record linkage, участь у написанні рукопису; Федоренко З.П., Гулак Л.О., Лапікура О.В. – участь у написанні рукопису; Тронько М.Д. – загальне керівництво, редагування висновків.

Фінансування: дослідження проведено при фінансовій підтримці Українсько-Американської програми «Науковий проєкт вивчення раку та інших захворювань щитоподібної залози в Україні в результаті аварії на Чорнобильській АЕС».

Декларація з етики: автори задекларували відсутність конфлікту інтересів і фінансових зобов'язань.

Стаття: надійшла до редакції 05.01.2026р.; перероблена 15.01.2026р.; прийнята до друку 09.04.2026 р.; надрукована 20.04.2026 р.

For citation: Gorokh EL, Shpak VM, Sumkina OV, Ryzhov AYU, Gorokh EL, Shpak VM, Sumkina OV, Ryzhov AYU, Fedorenko ZP, Gulak LO, Lapikura OV, Tronko MD. Identification of cases of malignant tumors among participants of epidemiology studies of the long-term consequences of the Chornobyl accident Fedorenko ZP, Gulak LO, Lapikura OV, Tronko MD. Identification of cases of malignant tumors among participants of epidemiology studies of the long-term consequences of the Chornobyl accident. *Endokrynologia.* 2026;31(1):97-103. DOI: 10.31793/1680-1466.2026.31-1.97.

Correspondence address: Shpak Viktor Mykhailovych, v.m.shpak@gmail.net; SI «VP Komisarenko Institute of Endocrinology and Metabolism of the NAMS of Ukraine», Vyshgorodska Str., 69, Kyiv 04114, Ukraine.

Information about the authors: Gorokh Evgeniy Leonidovich, Ph. D., Senior Researcher of the Nonprofit Organization «National Cancer Institute», ORCID: 0009-0000-9063-5824; Shpak Viktor Mykhailovich, Senior Researcher of the Department of Medical Consequences of the Chornobyl Accident and International Relations, SI «V.P. Komisarenko Institute of Endocrinology and Metabolism of the NAMS of Ukraine», ORCID: 0000-0002-6983-5490; Sumkina Olena Volodymyrivna – Leading Engineer, Nonprofit Organization «National Cancer Institute», ORCID: 0000-0003-2052-9043, Ryzhov Anton Yuriyovych Ph.D. Associate Professor, Taras Shevchenko National University of Kyiv, ORCID: 0000-0003-4099-0742; Fedorenko Zoya Pavlivna, Ph.D., M.D., Nonprofit Organization «National Cancer Institute», ORCID: 0000-0003-2322-9695; Gulak Lyudmila Olehivna, Ph.D., Leading Engineer of the Department of Medical Consequences of the Chornobyl Accident and International Relations, SI «V.P. Komisarenko Institute of Endocrinology and Metabolism of the NAMS of Ukraine», ORCID: 0009-0009-2086-0335; Lapikura Oles' Valeriyovych, Data Coordination Center Collaborator, ORCID: 0000-0001-7629-2933; Tronko Mykola Dmytrovych, MD, Ph.D., Prof., Corresponding Member NAS of Ukraine, Acad. NAMS of Ukraine, Head of the Department of Fundamental and Applied Problems of Endocrinology and Director of the SI «V.P. Komisarenko Institute of Endocrinology and Metabolism of the NAMS of Ukraine», ORCID: 0000-0001-7421-0981.

Personal contribution: Gorokh E.L. – idea, general management, research design, analysis of results, writing, design, preparation for publication of the paper; Shpak V.M. – participation in the development of the paper concept, participation in writing the manuscript; Sumkina O.V. – NCRU data sampling, participation in writing the manuscript; Ryzhov A.Yu. – development of record linkage tools, participation in writing the manuscript; Fedorenko Z.P., Gulak L.O., Lapikura O.V. – participation in writing the manuscript; Tronko M.D. – general management, discussion of the results obtained.

Funding: the study was carried out within the funding support of Ukrainian-American Program «Scientific Project to Study Cancer and Other Thyroid Diseases in Ukraine as a Result of the Chornobyl Nuclear Power Plant Accident».

Declaration of ethics: the authors have declared no conflicts of interest or financial obligations.

The article: received 05 January 2026; revised 15 January 2026; accepted April 09, 2026; published April 20, 2026.