

Аналіз поширення діабету в США: демографія, поведінкові та клінічні фактори



Вступ

Цукровий діабет є одним із найпоширеніших хронічних захворювань у світі та становить серйозну загрозу для здоров'я населення через високий ризик ускладнень і супутніх станів. Актуальність вивчення діабету зросла особливо після пандемії COVID-19, яка могла спричинити зміни у поведінкових факторах, способі життя та виявленні захворювання. Зростання поширеності діабету потребує глибшого аналізу як медичних, так і соціальних факторів ризику.

Цей звіт присвячений аналізу даних щодо захворюваності на діабет серед населення США. Метою дослідження є виявлення ключових демографічних, поведінкових та клінічних факторів, що пов'язані з діабетом, а також візуалізація та інтерпретація виявлених закономірностей.

Дослідження базується на наборі даних, що містить інформацію про понад 100 000 пацієнтів різного віку, статі, раси, а також включає показники ІМТ, рівня глюкози, HbA1c, історію куріння, наявність гіпертензії та серцево-судинних захворювань. У вибірці представлені дані про 8 500 пацієнтів із діабетом та 91 500 здорових осіб. Аналіз проводився з урахуванням обох груп для виявлення критичних відмінностей і потенційних маркерів ризику.

Методологія

- **Обробка даних** проводилася у середовищі Python із використанням бібліотек pandas, numpy, matplotlib, seaborn.
- **Попередня обробка** включала створення категоріальних змінних (вікові групи), об'єднання значень змінної smoking_history у ширші категорії для спрощення аналізу.

- **Візуалізація** здійснювалася за допомогою Python (matplotlib, seaborn) та Tableau. У Tableau створено інтерактивний дашборд із можливістю фільтрації за віком, статтю, статусом діабету та іншими змінними.
- **Машинне навчання** не включено до основного звіту, але реалізовано в окремому Jupyter Notebook. Код моделювання, балансування класів (SMOTE), побудови моделей (RandomForestClassifier) та оцінки їх якості доступний у GitHub-репозиторії (посилання надається окремо).

Ключові висновки дослідження (Summary)

- **Вік, ІМТ та клінічні показники.** Пацієнти з діабетом у середньому старші, мають вищий ІМТ, рівень глюкози в крові та HbA1c, що вказує на наявність метаболічних порушень.
- **Зв'язок із віком.** Частота діабету, гіпертензії та серцевих захворювань зростає з віком, особливо після 40 років.
- **Коморбідність.** Діабет часто поєднується з гіпертензією та серцево-судинними хворобами, що ускладнює перебіг захворювання.
- **Фактор куріння.** Понад 40% пацієнтів з діабетом мають або мали досвід куріння, що підкреслює його значення як фактора ризику.
- **Динаміка у часі.** У 2020–2022 роках зафіксовано погіршення клінічних показників у діабетиків, ймовірно, через вплив пандемії COVID-19.
- **Необхідність комплексного підходу.** Профілактика та лікування діабету мають охоплювати не лише контроль глюкози, а й управління супутніми захворюваннями та поведінковими факторами.

Епідеміологія діабету в США: динаміка, демографія, фактори поведінки та географія

Географічний розподіл здорових та хворих на діабет по штатах США

У наборі даних представлені пацієнти з усіх штатів США та деяких територій. Кількість пацієнтів у кожному штаті варіюється, але загалом розподіл є доволі рівномірним для більшості штатів із незначними відхиленнями. **Найбільша кількість** пацієнтів зареєстрована у штатах **Kentucky, Iowa, Hawaii та Nebraska** - по **2038** пацієнтів у кожному. **Найменша кількість** пацієнтів спостерігається у штатах **Wisconsin та Wyoming** - лише по **388** пацієнтів. Серед територій найменш представлені Virgin Islands (763 пацієнти) та Guam (1204 пацієнти).

Примітка: частина записів містить локацію як "United States", що ймовірно означає відсутність конкретної інформації про місце проживання або загальне позначення (всього 1401 запис).

Найвищі показники кількості пацієнтів **із діабетом** спостерігаються у штатах:

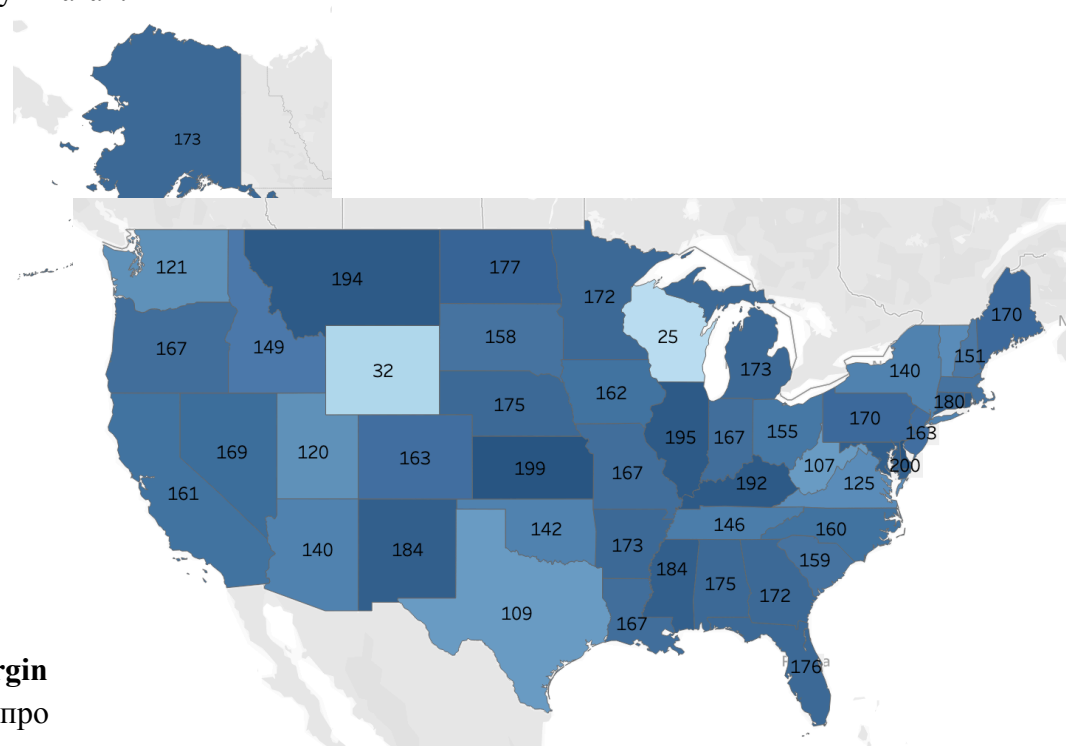
- **Delaware** - 200 пацієнтів
- **Kansas** - 199
- **Hawaii** - 186
- **Illinois** - 195
- **Montana** - 194
- **Kentucky та Rhode Island** - по 192

Найнижчі значення зареєстровані у штатах:

- **Wisconsin** - 25 пацієнтів
- **Wyoming** - 32
- **Virgin Islands** - 72
- **West Virginia** - 107
- **Texas** - 109 (несподівано низьке значення для великого штату)

Значення для деяких територій (наприклад, **Puerto Rico, Guam, Virgin Islands**) та об'єднаної категорії **United States** можуть свідчити про неповну або узагальнену інформацію про місце проживання.

Дані свідчать про нерівномірний розподіл пацієнтів із діабетом по території США, що може впливати на регіональні особливості доступу до медичних послуг та ефективності профілактичних програм.



Динаміка захворюваності на діабет за роками

Аналіз динаміки захворюваності на діабет у період з 2015 по 2022 рік виявив **різке зростання частки хворих у 2021–2022 роках**. Якщо протягом 2015–2019 років показник залишався стабільним на рівні близько **9%**, то у 2020 році відбулося тимчасове зниження до **7%**, за яким послідувало **різке зростання до 15% у 2021 році і понад 25% у 2022 році**.

Це зростання, ймовірно, є **наслідком пандемії COVID-19**, яка могла вплинути на:

- **Метаболічне здоров'я пацієнтів**: вірус може порушувати толерантність до глюкози, підвищувати рівень HbA1c, а також провокувати нові випадки діабету навіть у людей без попередньої історії захворювання.
- **Функцію β-клітин підшлункової залози**, через запалення та стресову відповідь організму під час тяжкого перебігу COVID-19.
- **Стероїдну терапію**, що застосовувалась при лікуванні COVID-19 і могла посилювати інсулінорезистентність.
- **Поведінкові фактори**, зокрема зниження фізичної активності, порушення харчування та підвищений рівень стресу внаслідок ізоляції.

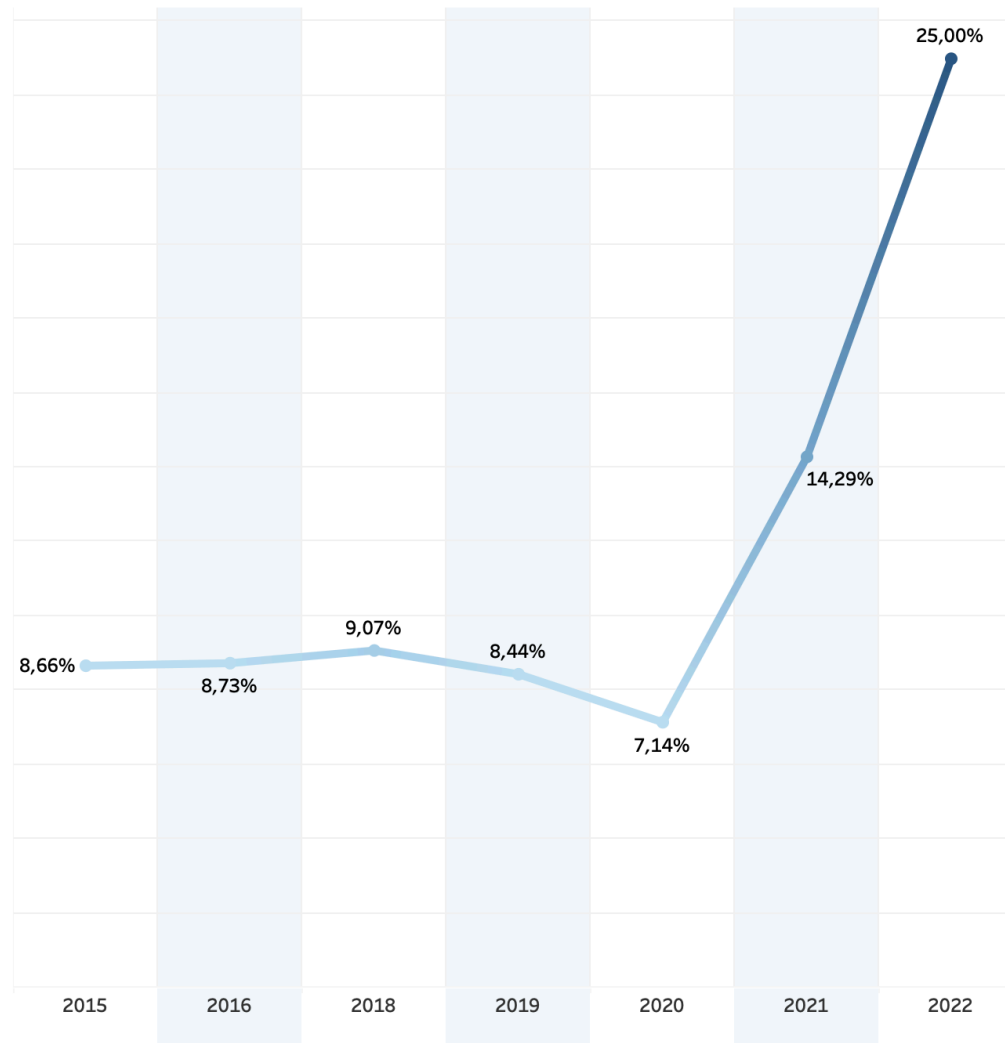
Ці фактори, разом із можливими структурними змінами у вибірці або кращою діагностикою, ймовірно, зумовили значне зростання кількості нових випадків діабету в останні роки.

Цей тренд відповідає спостереженням CDC (США) та висновкам, опублікованим у **Nature Reviews Endocrinology**, щодо підвищеного ризику виникнення діабету протягом року після перенесеного COVID-19, особливо серед дорослого населення.

Посилання на джерело:

https://pmc.ncbi.nlm.nih.gov/articles/PMC8757617/?utm_source=chatgpt.com

Динаміка захворюваності на діабет (%), 2015–2022

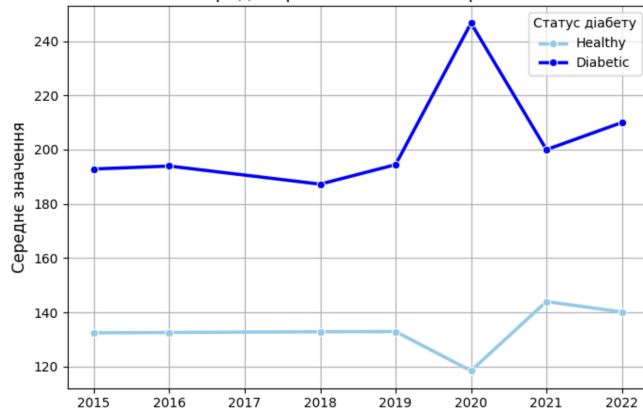


Динаміка середніх клінічних показників пацієнтів (2015–2022)

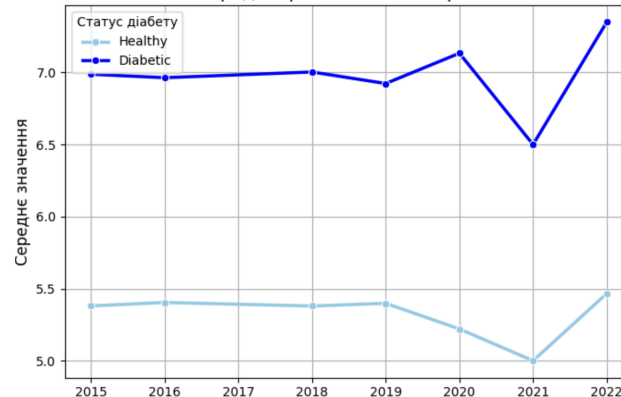
Аналіз середніх клінічних показників за період 2015–2022 років демонструє **стійкі відмінності між пацієнтами з діабетом і тими, хто не має цього діагнозу**. Упродовж усього періоду в осіб із діабетом фіксується **значно вищий середній вік, індекс маси тіла (ІМТ), рівень глікемії та HbA1c**, що свідчить про **хронічні метаболічні порушення, притаманні цій групі**.

Особливо показовими є зміни, зафіксовані в **2020–2022 роках**: серед пацієнтів із діабетом спостерігається **помітне зростання рівня глюкози та глікованого гемоглобіну**. Це може свідчити про зниження ефективності контролю за захворюванням під впливом зовнішніх факторів, зокрема пандемії COVID-19, яка суттєво вплинула на доступ до медичних послуг, режим харчування, фізичну активність та рівень стресу.

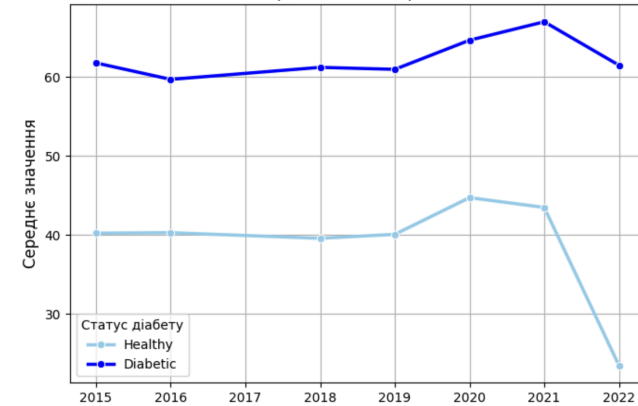
Середній рівень глюкози по роках



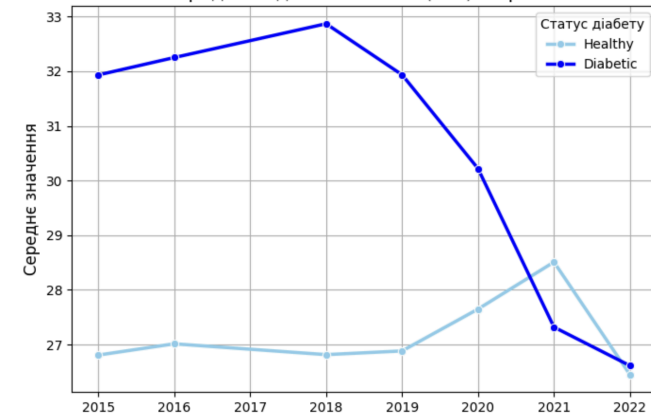
Середній рівень HbA1c по роках



Середній вік по роках



Середній індекс маси тіла (BMI) по роках



Для здорових пацієнтів загальна динаміка показників є більш стабільною. Водночас у **2020–2021 роках також зафіксовано помірне зростання середнього віку та ІМТ**, що може відображати загальнонаціональні тенденції - такі як малорухливий спосіб життя, збільшення маси тіла та інші наслідки ізоляції й карантинних обмежень.

Загалом, дані підкреслюють **важливість безперервного моніторингу клінічних показників, особливо в умовах соціальних або епідеміологічних криз**, а також необхідність посилення профілактичної медицини в періоди підвищеного ризику для вразливих груп населення.

Гендерний розподіл серед пацієнтів з діабетом

Аналіз показує, що серед усієї вибірки:

- Жінки становлять більшість як серед здорових осіб (59.1%), так і серед пацієнтів із діабетом (52.5%).
- Проте частка чоловіків серед хворих на діабет є помітно вищою (47.5%), ніж серед здорових (40.9%).

Це може вказувати на вищий ризик розвитку діабету серед чоловіків у порівнянні з жінками, або на структурні відмінності в поведінкових і фізіологічних факторах між статтями (наприклад, маса тіла, рівень активності, стиль життя тощо).

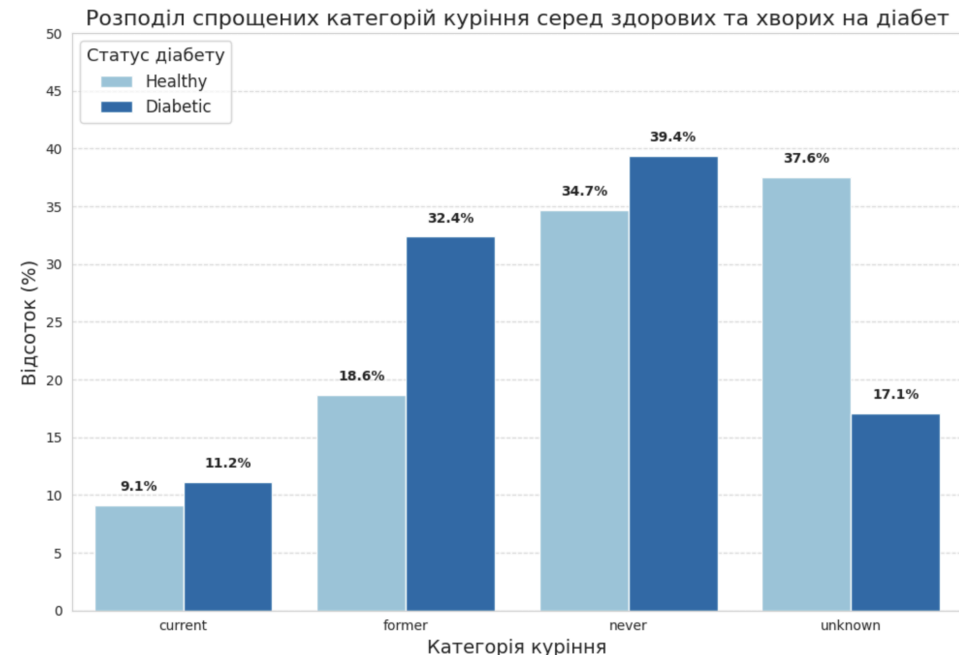
Розподіл категорій куріння серед здорових та хворих на діабет

Аналіз розподілу категорій куріння серед здорових пацієнтів і пацієнтів із діабетом виявив значущі відмінності:

- Серед здорових найбільшу частку становлять особи з невідомим статусом куріння (37.6%) та ті, хто ніколи не курив (34.7%). Частка поточних курців - 9.1%, колишніх - 8.5%.
- Серед хворих на діабет переважають особи, які ніколи не курили (39.4%), проте сукупна частка нинішніх, колишніх і періодичних курців перевищує 43%:
 - Former - 18.7%
 - Current - 11.2%
 - Ever + Not current - $\approx 13.7\%$

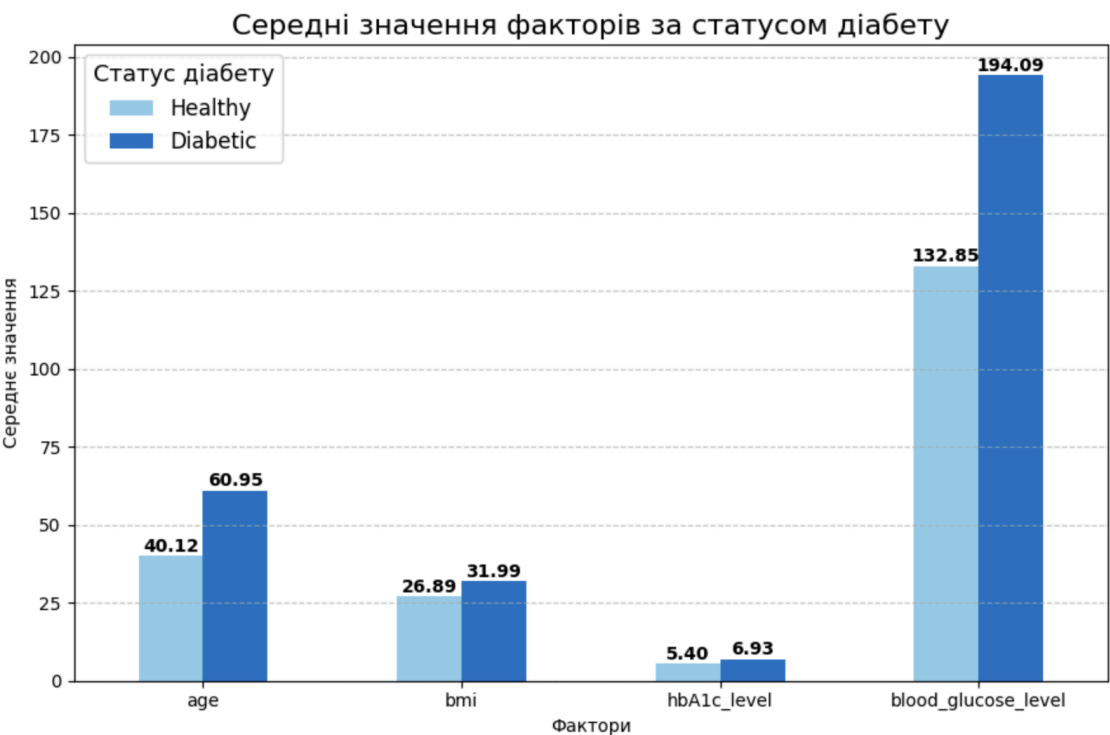
Це свідчить про значно вищу поширеність куріння серед осіб із діабетом, що може вказувати на потенційний зв'язок між курінням і розвитком діабету, або на зміну поведінки після встановлення діагнозу.

Також важливо зазначити, що в 17.1% випадків інформація про куріння відсутня, що свідчить про потребу в покращенні збору анамнестичних даних у медичній практиці.



Середні значення рівня глюкози, віку, BMI, HbA1c

Аналіз середніх фізіологічних показників виявив **чіткі відмінності між пацієнтами з діабетом і без нього**:



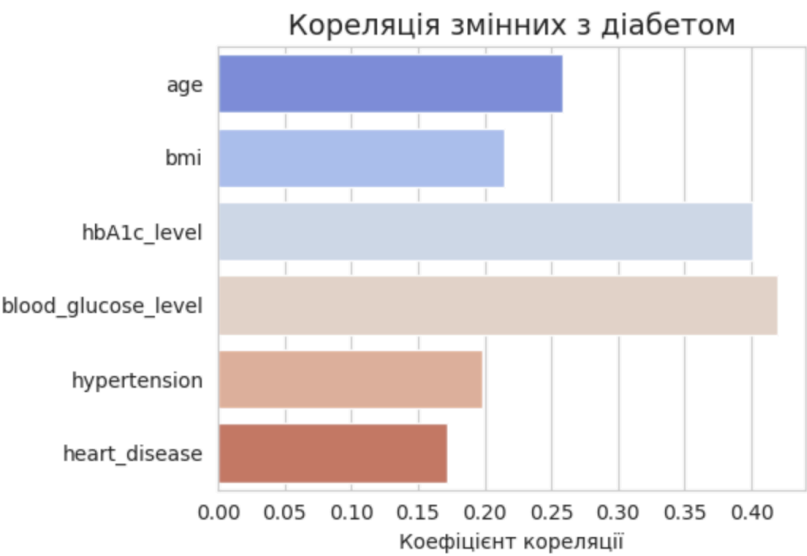
- **Середній вік** пацієнтів із діабетом становить **60.9 року**, що на понад **20 років більше**, ніж у здорових осіб (40.1 року). Це може свідчити про **вікову природу ризику розвитку діабету**.
- **Індекс маси тіла (BMI)** у людей з діабетом - **31.99**, що відповідає **ожирінню**. У здорових осіб він становить **26.89**, тобто **надмірну вагу**.
- **Рівень HbA1c** - середній показник цукру за 2–3 місяці - значно вищий у діабетиків: **6.93% проти 5.40%**.
- **Середній рівень глюкози в крові** також вищий: **194.1 мг/дл у пацієнтів із діабетом проти 132.9 мг/дл у здорових**.

Ці результати узгоджуються з клінічними критеріями діагностики діабету та підтверджують, що **надмірна вага, вік та гіперглікемія є ключовими факторами, пов'язаними з цим захворюванням**.

Кореляція змінних з наявністю діабету

Найсильніший позитивний кореляційний зв'язок спостерігається між діабетом та **рівнем глюкози в крові** ($r \approx 0.42$) та **рівнем HbA1c** ($r \approx 0.40$), що логічно, оскільки ці показники є **ключовими маркерами цукрового діабету**. **Вік** також має **помірний зв'язок з діабетом** ($r \approx 0.26$), що свідчить про збільшення ризику із зростанням віку.

Індекс маси тіла, гіпертензія та серцеві захворювання також мають **помірний позитивний зв'язок**, що підтверджує роль метаболічних і серцево-судинних факторів у розвитку діабету.



Розподіл пацієнтів за історією куріння

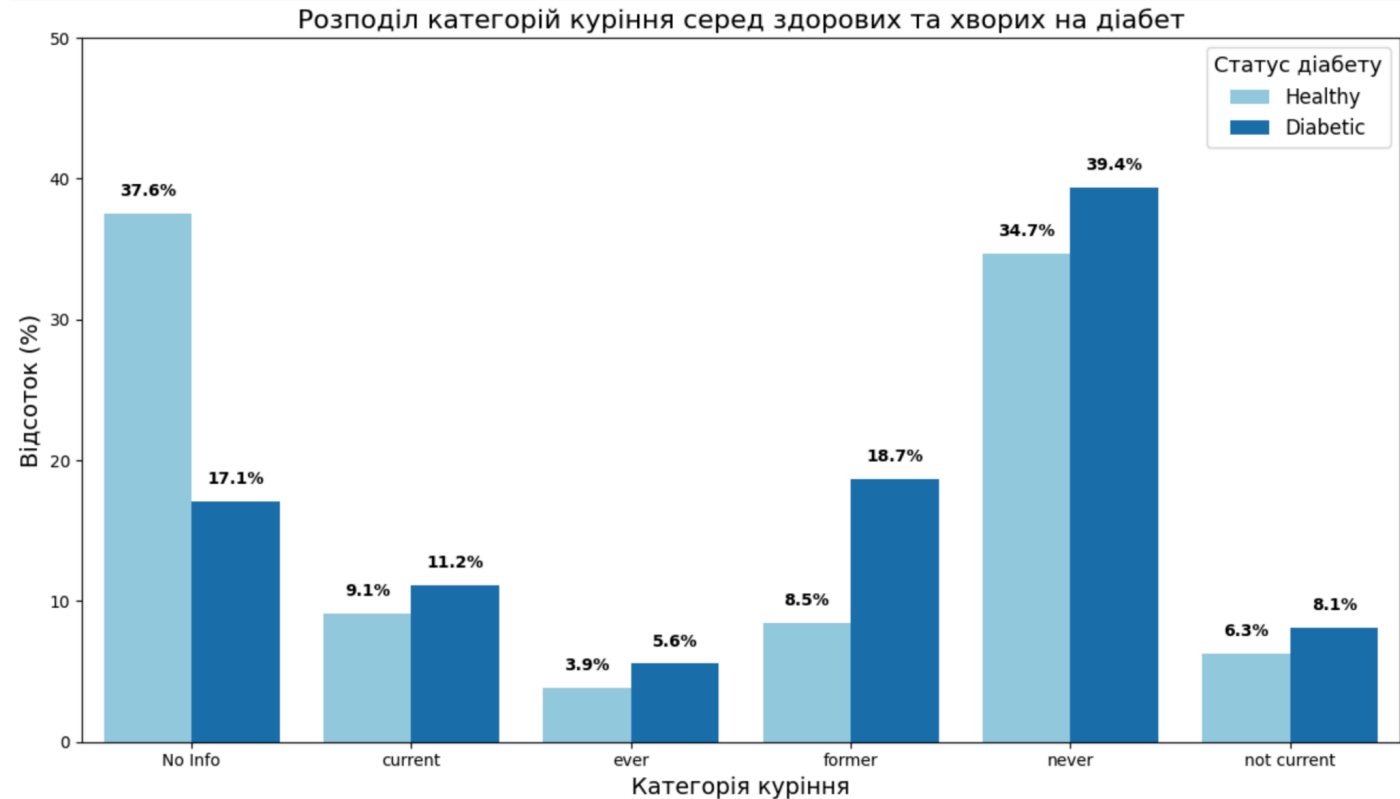
Аналіз розподілу історії куріння серед здорових пацієнтів та пацієнтів із діабетом демонструє, що серед діабетиків частка нинішніх і колишніх курців суттєво вища. Понад 40% пацієнтів із діабетом мають або мали досвід куріння, що підкреслює важливість врахування куріння як фактора ризику при профілактиці та лікуванні діабету.

Аналіз розподілу історії куріння серед здорових пацієнтів показує, що **найбільша частка** припадає на осіб з **невідомим статусом куріння** ("No Info") - 37.6% та тих, хто **ніколи не курив** ("Never") - 34.7%. **Поточні курці становлять 9.1%**, а колишні - 8.5%, при цьому всі категорії, пов'язані з курінням, мають нижчі показники, ніж у пацієнтів з діабетом.

У групі пацієнтів із діабетом **найбільша частка належить тим, хто ніколи не курив - 39.4%**. Водночас **сукупна частка осіб із курильним досвідом є значною**: колишні курці становлять 18.7%, поточні - 11.2%, а інші категорії куріння охоплюють близько 13.7%. Таким чином, понад 40% пацієнтів із діабетом мають або мали досвід куріння, що підкреслює **важливість врахування куріння як фактора ризику** при профілактиці та лікуванні цього захворювання. **Водночас 17.1% даних про куріння відсутні**, що свідчить про необхідність поліпшення збору анамнестичної інформації у клінічній практиці.

Загалом, підвищена частка нинішніх та колишніх курців серед хворих на діабет може свідчити як про потенційний зв'язок між курінням і розвитком діабету, так і про можливу зміну курильних звичок після встановлення діагнозу.

Посилання на статтю: <https://pubmed.ncbi.nlm.nih.gov/18073361/>



Клінічні та поведінкові показники за віковими групами

Середні значення BMI, глікемії, HbA1c, розподіл куріння, частка пацієнтів з гіпертензією та серцевими захворюваннями

Аналіз частки пацієнтів із діабетом за віковими групами виявляє чітку вікову залежність захворюваності:

- У молодших вікових групах (до **30 років**) частка діабетиків залишається низькою — від **1% до 5%**.
- Починаючи з **30–39 років**, спостерігається поступове **зростання** захворюваності, що досягає піку у віковій групі **50–59 років** — понад **26%**.
- Після **60 років** рівень діабету **знижується**, однак залишається високим: близько **19.4%** у групі **60–69 років** та **12%** у віковій категорії **70–79 років**.

Такий розподіл свідчить про **накопичувальний ефект метаболічних та поведінкових факторів з віком**, з подальшим зниженням кількості діабетиків у старшому віці, ймовірно, через підвищену смертність серед хворих або зменшення частки представників цієї вікової групи у вибірці.

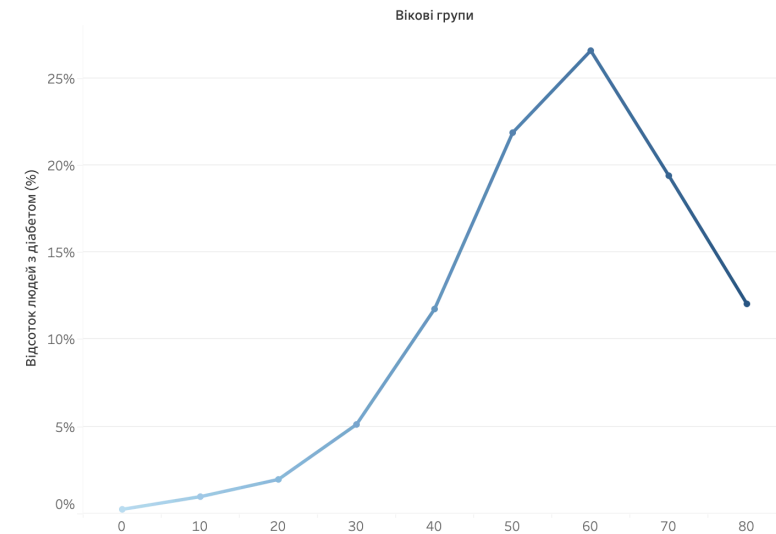
Динаміка індексу маси тіла (ІМТ) за віком

Аналіз ІМТ за віковими групами пацієнтів з діабетом демонструє чітку вікову тенденцію до зростання маси тіла у молодому та середньому віці з подальшим невеликим зниженням у старших групах:

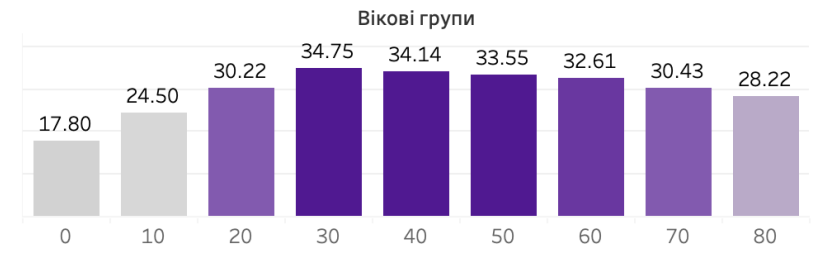
- У дітей та підлітків (**0–9 років**) **ІМТ - 17.8**, що нижче норми (можливо, пов'язано з пропорціями росту).
- У віці **10–19 років** **ІМТ піднімається до 24.5** - межа норми.
- У дорослому віці (20–59 років) фіксується стабільне ожиріння: ІМТ коливається від 30.2 до 34.75, що свідчить про високий ризик метаболічних порушень.
- Після **60 років** **ІМТ поступово зменшується, досягаючи 28.2** у віці 80+, але залишається в зоні надмірної ваги.

Такий розподіл підтверджує, що пік маси тіла припадає на середній вік, що збігається з віковими показниками діабету, і може бути ключовим фактором ризику.

Ризик захворюваності на діабет за віковими групами



Середній показник BMI серед пацієнтів з діабетом



Динаміка HbA1c за віком

Це один із ключових показників контролю глюкози в організмі за останні 2–3 місяці.

Нормальні значення:

- < 5.7% - норма
- 5.7–6.4% - предіабет
- ≥ 6.5% - діабет

Значення глікованого гемоглобіну (HbA1c) стабільно **перевищують діабетичний поріг (≥6.5%) у всіх вікових групах.**

Найвищий показник - у наймолодшій групі (**0–9 років**): **7.18%**, що може свідчити про наявність цукрового діабету 1 типу або специфіку вибірки в цій категорії.

У дорослих віком **10–80 років рівень HbA1c тримається на стабільному рівні 6.89–6.96%**, без чіткої тенденції до зростання чи зниження з віком. Це вказує на однаково високий рівень хронічної гіперглікемії серед усіх вікових категорій пацієнтів з діабетом.

Така стабільність може бути наслідком тривалого перебігу хвороби, неоптимального контролю глюкози або однакових стратегій лікування незалежно від віку.

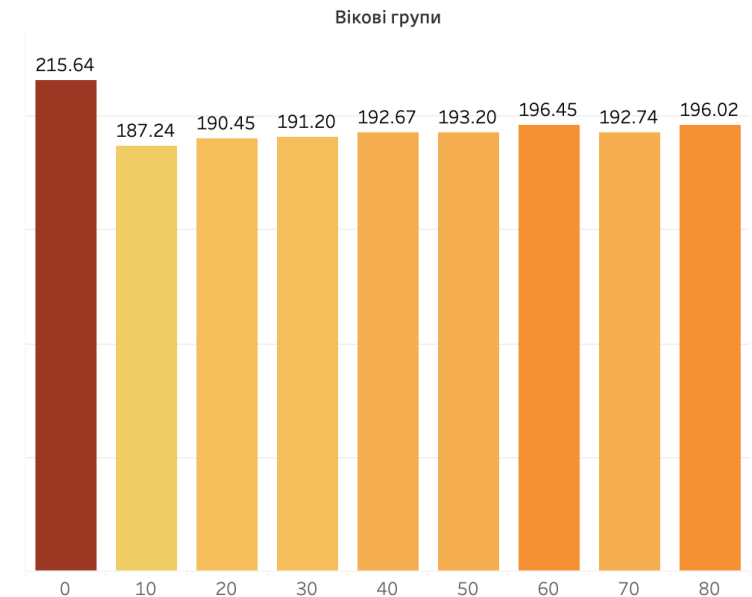
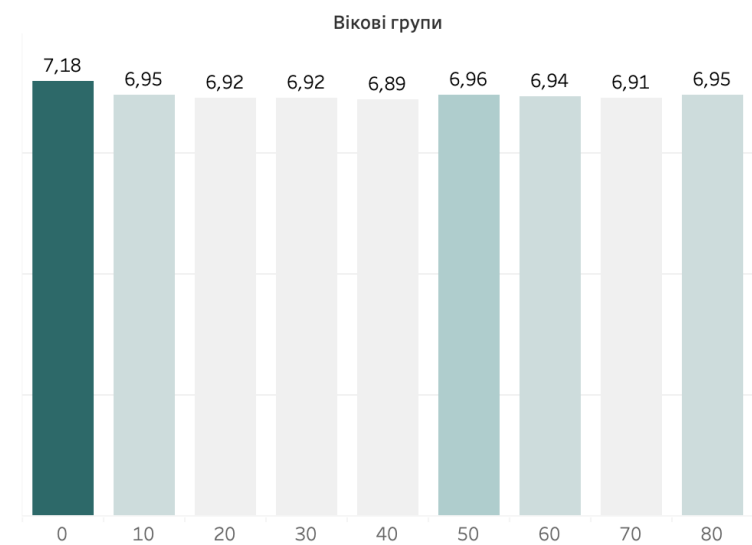
Рівень глюкози за віком

Рівень глюкози в крові **перевищує діабетичний поріг (≥126 mg/dL натще) в усіх вікових групах**, що підтверджує наявність вираженої гіперглікемії серед пацієнтів з діабетом. Найвищі показники спостерігаються у: наймолодшій групі (**0–9 років**): **215.6 mg/dL**, людей похилого віку (**60–69 та 80+**): **≈196 mg/dL**. Натомість у групах **10–59 років** рівень глюкози залишається на дещо нижчому, але стабільно високому рівні (**~187–193 mg/dL**).

Така динаміка може свідчити про:

- Складнощі в контролі діабету серед наймолодших і найстарших пацієнтів
- Зв'язок між віком і компенсацією захворювання
- Можливо, меншу ефективність терапії або відсутність своєчасної корекції лікування у крайніх вікових групах.

Середній показник глікованого гемоглобіну HbA1c



Відсоток пацієнтів з гіпертензією за віковими групами

Аналіз поширеності гіпертензії серед пацієнтів у різних вікових групах показує чітку тенденцію до збільшення з віком.

У наймолодших категоріях (**0-20 років**) відсоток пацієнтів з гіпертензією **практично відсутній** або дуже низький (**0-7,14%**). Від **30 років і старше** спостерігається поступове зростання частоти гіпертензії: від **14,9%** у віковій групі **30-39 років** до максимальних показників близько **28%** у вікових категоріях старших **60 років**.

Після досягнення піку у віковій групі **70-79 років (28,08%)** відсоток пацієнтів з гіпертензією стабілізується або дещо знижується (**27,73% у 80+**).

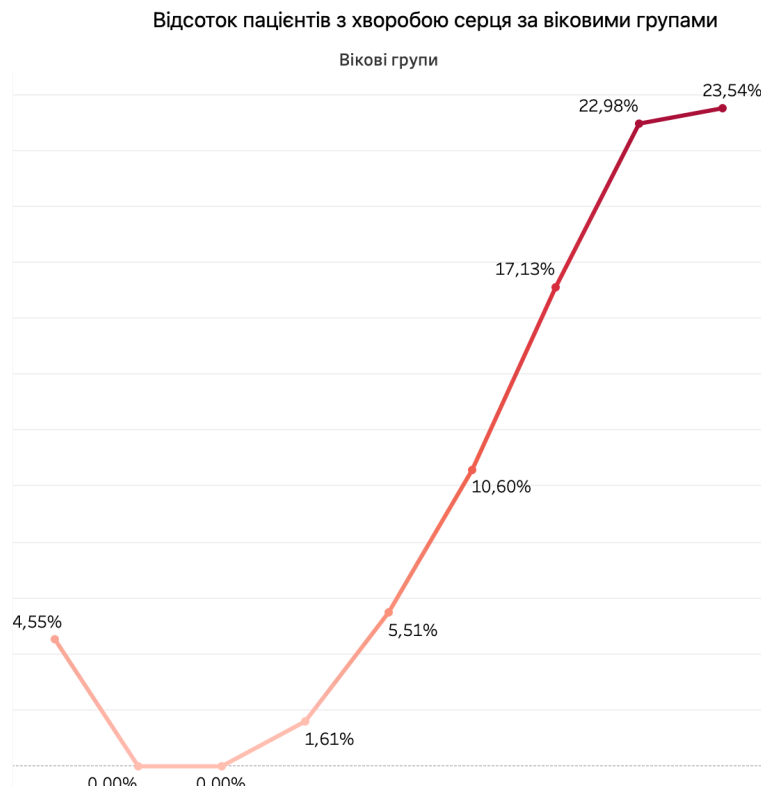
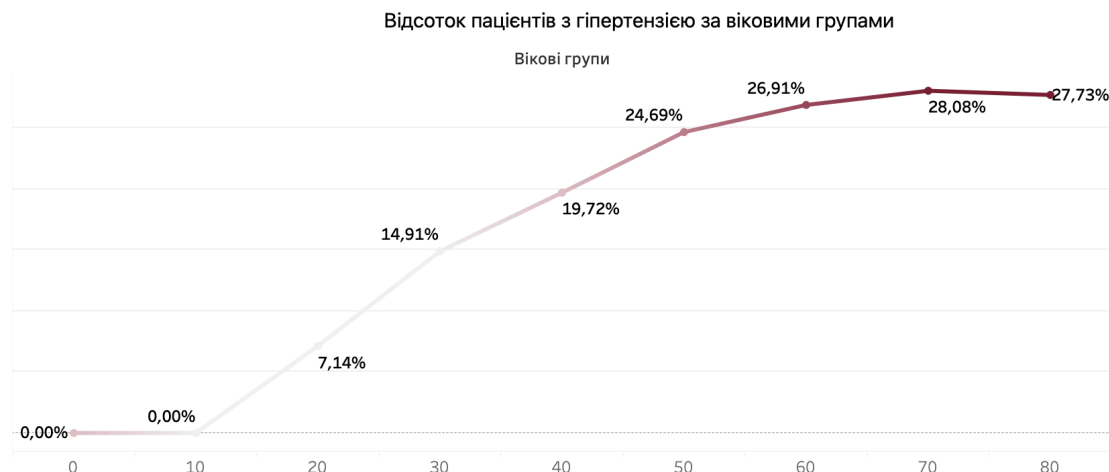
Ці дані свідчать про тісний зв'язок гіпертензії з віком, що відповідає світовим клінічним спостереженням і підкреслює необхідність більш пильного моніторингу та профілактики артеріальної гіпертензії у старших вікових групах.

Розподілу серцевих захворювань за віковими групами

Аналіз поширеності серцевих захворювань у різних вікових категоріях демонструє чітку вікову залежність.

У наймолодших групах (**0-20 років**) захворюваність практично відсутня або дуже низька (**0-4,55%**). Починаючи з **30-39 років**, частка пацієнтів із серцевими захворюваннями починає зростати — від **1,61%** до **5,51%**, і надалі помітно збільшується у старших вікових групах: **10,6%** (40-49 років), **17,13%** (50-59 років), **22,98%** (60-69 років) та досягає максимуму **23,54%** у пацієнтів віком **80 років і старше**.

Ця тенденція свідчить про посилення ризику серцево-судинних захворювань із віком, що відповідає загальноприйнятим медичним даним і підкреслює важливість регулярного кардіологічного обстеження у старших вікових категоріях для раннього виявлення та профілактики ускладнень. Посилання на статтю: <https://pubmed.ncbi.nlm.nih.gov/8452139/>

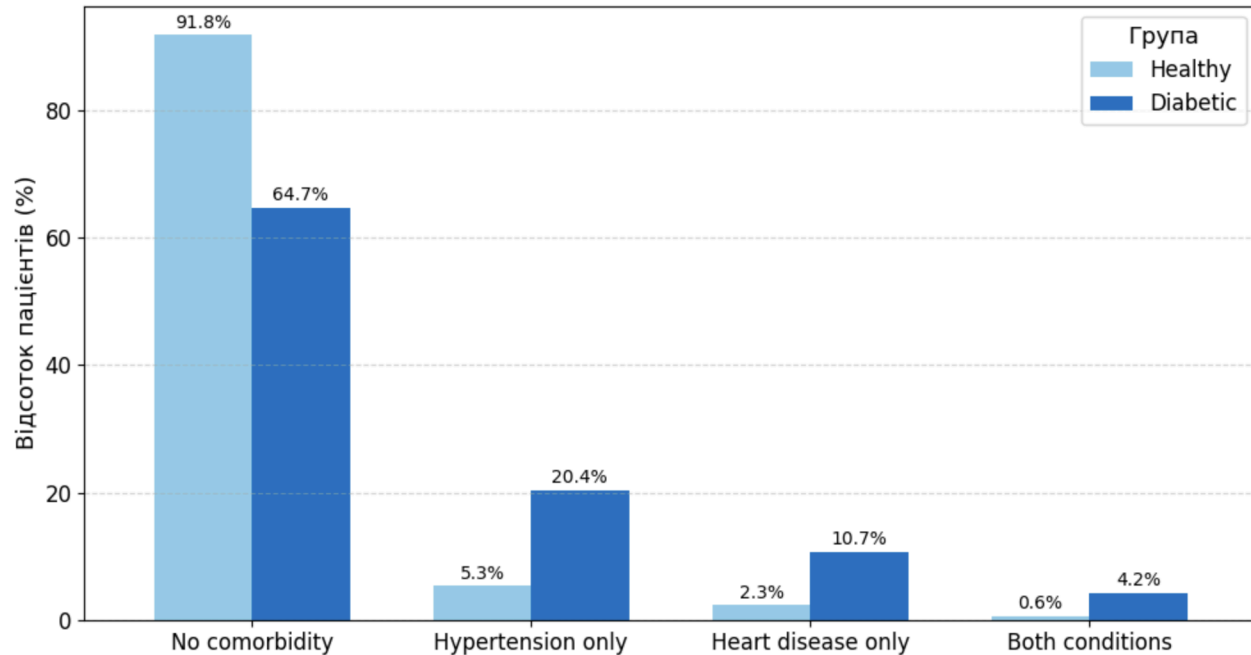


Супутні захворювання серед пацієнтів з діабетом та без

Аналіз поширеності гіпертензії та серцевих захворювань серед здорових та пацієнтів з діабетом демонструє суттєві відмінності:

- 91.8% осіб без діабету не мають жодної з цих патологій, тоді як серед пацієнтів із діабетом таких лише 64.7%.
- Гіпертензія без супутньої серцевої патології спостерігається у 5.3% здорових проти 20.4% людей з діабетом..
- Серцеві захворювання без гіпертензії — у 2.3% здорових та 10.7% осіб із діабетом.
- Поєднання гіпертензії та серцевої патології виявлено у 0.6% здорових і у 4.2% діабетичних пацієнтів.

Порівняння коморбідності у здорових та діабетиків



Висока частота супутніх станів у групі пацієнтів з діабетом підтверджує її **більшу клінічну вразливість та складність ведення**. Ці результати вказують на необхідність міждисциплінарного підходу до таких пацієнтів, враховуючи не лише контроль глікемії, а й ретельний моніторинг серцево-судинного стану та артеріального тиску.

Посилання на статтю: <https://pubmed.ncbi.nlm.nih.gov/33651559/>, <https://pubmed.ncbi.nlm.nih.gov/29884191/>

Загальні висновки

1. Вікові та клінічні відмінності між пацієнтами з діабетом та без нього

Пацієнти з діабетом суттєво старші за здорових осіб, мають вищий індекс маси тіла (ІМТ), рівні глюкози в крові та глікованого гемоглобіну (HbA1c). Це підтверджує наявність у них виражених метаболічних порушень і підвищеного ризику ускладнень.

2. Розподіл захворюваності за віковими групами

Частота діабету зростає з віком, особливо після 40-50 років, що узгоджується з відомими епідеміологічними даними. Аналогічно, з віком збільшується поширеність гіпертензії та серцево-судинних захворювань.

3. Рівень коморбідності

У пацієнтів з діабетом значно вища частота гіпертензії та серцевих хвороб у порівнянні зі здоровими особами. Особливо помітно, що у діабетиків частіше зустрічається поєднання цих захворювань, що ускладнює перебіг основної хвороби та вимагає комплексного підходу до лікування.

4. Вплив куріння

Понад 40% пацієнтів з діабетом мають або мали досвід куріння, що свідчить про значний вплив цього фактора на розвиток і перебіг діабету. Виявлена висока частка відсутніх даних щодо куріння вказує на потребу покращення збору анамнезу.

5. Тенденції динаміки клінічних показників у часі (2015–2022)

Протягом останніх років у пацієнтів з діабетом спостерігається зростання рівнів глюкози та HbA1c, що може бути пов'язано з погіршенням контролю захворювання, зокрема через вплив пандемії COVID-19. У здорових пацієнтів зміни менш виражені, хоча й фіксується поступове збільшення віку та ІМТ.

6. Рекомендації

Враховуючи високий рівень коморбідності та вплив факторів способу життя (куріння, ІМТ), необхідні комплексні профілактичні та лікувальні заходи, які враховують не тільки контроль глюкози, а й управління супутніми захворюваннями і модифікацію факторів ризику. Важливо також покращити збір анамнестичних даних для більш точного моніторингу пацієнтів.

Рекомендації для подальших спостережень та досліджень

На основі отриманих результатів доцільно розширити перелік досліджуваних змінних у майбутніх дослідженнях і включити такі показники:

1. Біомаркери для раннього виявлення діабету та метаболічних порушень:

- Рівень інсуліну та С-пептиду — для оцінки функції β -клітин та інсулінорезистентності.
- Глікований альбумін — як альтернативний короткостроковий маркер глікемії (1–3 тижні).
- Показники запалення:
 - високочутливий С-реактивний білок (hs-CRP),
 - інтерлейкін-6 (IL-6),
 - фактор некрозу пухлин- α (TNF- α) — хронічне системне запалення відіграє ключову роль у розвитку діабету.
- Адипонектин — знижений рівень пов'язаний з інсулінорезистентністю.
- Показники оксидативного стресу, зокрема:
 - перекис водню (H_2O_2),
 - малоновий діальдегід (MDA),
 - 8-ізопростани — ці маркери можуть сигналізувати про ушкодження клітин на доклінічному етапі.

2. Оцінка ускладнень та супутніх станів:

- Мікроальбумінурія — ранній маркер діабетичної нефропатії.
- Обстеження органів-мішеней:
 - Офтальмологічне обстеження (стан судин сітківки) — для виявлення діабетичної ретинопатії.
 - Кардіо-метаболічні показники — включаючи холестерин, тригліцериди, артеріальний тиск.

3. Соціально-поведінкові чинники:

- Рівень фізичної активності — малорухливий спосіб життя підвищує ризик діабету.
- Професія / умови праці — для оцінки впливу стресу, змінного графіку, шкідливих звичок тощо.
- Сімейний статус / наявність дітей — як потенційний фактор стресу або соціальної підтримки.