Оцінювання хірургічних доказів

1

Томас М. Дрейк (Thomas M. Drake), Юен М. Гаррісон (Ewen M. Harrison)

ВСТУП ДО ХІРУРГІЧНИХ ДОКАЗІВ

Відомо, що доказова медицина є відносно недавнім нововведенням. Лише протягом останніх двох століть наукові методи стали загальноприйнятим засобом визначення найефективніших методів лікування та тестів. Проте ніде в медицині ця трансформація не досягла такого розвитку, як у хірургічних дисциплінах, де численні інновації проклали шлях до сучасних методів лікування.

Отримання високоякісних доказів у хірургії може бути особливо складним. Причин, що лежать в основі цього, принаймні чотири. По-перше, сама операція є складним втручанням. До того ж існує багато змінних, які необхідно враховувати при плануванні обстежень та післяопераційного догляду, опрацюванні великої різноманітності хірургічних методик та вивченні природних явищ, необхідних хірургам для освоєння нових підходів або оперативних втручань.

По-друге, самі хірурги розділилися у своєму ставленні до досліджень. Опитування австралійських хірургів показало, що вони вірять в те, що їхня власна клінічна практика є кращою за клінічні настанови і що доказова хірургія негативно впливає на прийняття рішень для пацієнтів, а відмова від використання доказів у прийнятті клінічних рішень не впливає негативно на лікування пацієнтів. Таке ставлення є серйозною перешкодою на шляху впровадження доказової медицини в хірургію і потребує вирішення, якщо ми хочемо, щоб результати та догляд за пацієнтами були якнайкращими[1].

По-третє, багато операцій мають історичне походження, коли операція виконувалася за певними показаннями, фактично для встановлення діагнозу. Використання таких втручань має бути обмеженим, оскільки було б неетично відмовляти пацієнтам в ефективному і перевіреному лікуванні на основі сучасних методів дослідження.

Ну і нарешті, брак фінансування та зацікавленості призвів до того, що наукові дослідження стали швидше винятком, ніж нормою. На сьогодні незвичною є ситуація, коли пацієнт, якому проводять хірургічне втручання, стає учасником наукового дослідження. Однак, коли ці чинники розглядаються разом, то не дивно, що у 1998 р. редактор журналу «*Ланцет*» порівняв хірургічні дослідження з «комічною оперою»[2].

Упродовж останніх 20 років відбулися значні покращення. Кілька масштабних ініціатив, започаткованих хірургічними організаціями та співпрацею, привели до створення нових, специфічних для хірургії дослідницьких структур (наприклад, структура IDEAL https://www.ideal-collaboration.net). Це виявляється успішним, оскільки хірургічні дослідження все частіше публікуються в найбільших медичних журналах світу.

ЗМІНЮЄМО СВІТ ЗА ДОПОМОГОЮ ДОКАЗІВ

Для того щоб хірургічні дослідження змінили практику і покращили лікування пацієнтів, вони повинні стосуватися вирішення питання про область справжньої клінічної невизначеності. Невизначеність щодо того, яке лікування є найкращим, називається «клінічною рівновагою». Ключовим припущенням для етичного проведення будь-якого інтервенційного дослідження є наявність клінічної рівноваги, тобто повинна існувати справжня невизначеність щодо того, яке лікування є найкращим для даної групи пацієнтів. Якщо клінічної рівноваги немає і достеменно відомо, що одне лікування є кращим, неетично свідомо піддавати пацієнтів гіршому лікуванню.

З наведеними нижче дослідженнями можна ознайомитися, беручи до уваги таке:

- Хто входить до складу популяції пацієнтів або є цільовим станом (цільовою групою)?
- Яким було втручання (для інтервенційних досліджень, тобто клінічних випробувань) або вплив (для обсерваційних досліджень)?
- Якою була група порівняння (або контрольна група)?
- Як вимірювався ефект від втручання і чи був він виміряний точно?
- Чи існують джерела упередженості або введення в оману?
- Яким був результат?
- Чи є це дослідження актуальним для моєї клінічної практики?

У цьому розділі ми розглянемо кожен з цих пунктів більш детально[3–5].

MRC-CLASICC

У цьому дослідженні розглядалося, чи еквівалентна лапароскопічна хірургія раку товстої кишки відкритому хірургічному втручанню. Дослідження розпочалося в 1996 р., коли існувала справжня невизначеність щодо того, чи є лапароскопічна колоректальна хірургія безпечною та ефективною[3].

Дослідження показало, що при резекції товстої кишки з приводу раку лапароскопічне втручання настільки ж безпечне, як і відкрита операція. Авторами були опубліковані дані як про короткострокові, так і про довгострокові результати. У поєднанні ж із кількома іншими експериментами це дослідження проклало шлях до впровадження рутинної лапароскопічної колоректальної хірургії.

ФОРМУЛЮВАННЯ КЛІНІЧНОГО ПИТАННЯ

Дослідження STITCH

Дослідження «Техніка дрібних швів проти великих швів для закриття серединних розрізів черевної стінки» (STITCH) було подвійним сліпим, багатоцентровим, рандомізованим контрольованим дослідженням, опублікованим у 2015 р.[4] У ньому порівнювали два різні інтервали фасціального шва після закриття серединної лапаротомії — один вкол на 1 см проти одного вколу на 0,5 см.

Дослідження показало, що при використанні меншого інтервалу між швами під час закриття фасції частота виникнення защемлених гриж протягом першого року після операції знижується майже на 50 %. Це дослідження є чудовим прикладом того, як навіть невеликі технічні зміни можуть значно покращити результати лікування пацієнтів за умови належного оцінювання.

PROTECT

Це дослідження порівнювало хірургічне втручання з променевою терапією та спостереженням при локалізованому раку передміхурової залози[5]. Використання простатоспецифічного антигена (англ. *prostate-specific antigen*, PSA) для виявлення раку передміхурової залози є суперечливим, адже коли рівень PSA підвищений і виявлено рак, то стає неясно, чи є користь від лікування пухлин, виявлених таким чином.

Результати 10-річного дослідження PROTECT показали, що виживаність після лікування непоширеного раку передміхурової залози, виявленого за допомогою PSA, не відрізнялася між групами лікування. Встановлено, що хірургічне втручання та променева терапія мають нижчі показники прогресування захворювання, однак це пов'язано з ризиками. Результати цього дослідження створюють дилему для чоловіків з локалізованим раком передміхурової залози щодо необхідності вибору між спостереженням або радикальним втручанням, що може призвести до ускладнень з боку сечовидільної і статевої функцій, а також функції кишківника, оскільки обидва варіанти лікування мають схожі довгострокові результати виживання.

Першим кроком у розробці клінічного дослідження є формулювання гіпотези або питання дослідження. Розуміння складових частин клінічної гіпотези є ключовим для оцінювання актуальності та якості хірургічних досліджень.

Для формулювання клінічних питань можна використовувати простий, структурований підхід. Цей підхід враховує кілька важливих аспектів клінічного дослідження:

- «P» — Population — популяція, населення (пацієнти з цільовим захворюванням).
- «I» — Intervention or exposure — інтервенція або вплив (втручання або вплив, що вивчається).
- «C» — Comparison — порівняння (група порівняння, група з якою порівнюється втручання).
- «O» — Outcomes — результати (який був результат, що нас цікавить, і як він вимірювався).
- «S» — Study design — дизайн дослідження (як проводилось дослідження).

Такий підхід відомий як «PICOS». Ми можемо розглянути багато клінічних питань таким чином (блок 1.1).

Як приклад ми можемо розглянути застосування антибіотиків у дорослих з гострим апендицитом. У цій ситуації антибіотики застосовуються для зменшення ускладнень апендициту і зниження необхідності хірургічного втручання.

Використання підходу PICOS:

- P: популяція складається з дорослих із неперфорованим гострим апендицитом.
- I: втручання полягає в застосуванні антибіотиків.
- C: група порівняння — це група звичайного клінічного лікування (апендектомія).
- O: основним результатом є частота ускладнень, що визначена за допомогою валідованої системи оцінювання.
- S: дизайн дослідження є рандомізованим контрольованим дослідженням.

ПОПУЛЯЦІЯ (НАСЕЛЕННЯ) АБО ЦІЛЬОВИЙ СТАН

Досліджувана популяція — це група учасників, відібраних із загальної популяції на основі певних характеристик. Наявність чітко визначеної досліджуваної популяції є ключем до того, що дослідження є клінічно значущим і може дати відповідь на конкретне запитання. Якщо цільовий стан використовується як основа для відбору популяції, для визначення такої популяції слід застосовувати перевірені діагностичні критерії.

Підставою для включення пацієнтів у дослідження має бути систематичність, щоб уникнути упередженості. Найкращий спосіб переконатися в тому, що зразок є репрезентативним і вільним від упередженості — це підійти до кожного відповідного пацієнта послідовно. Часто це називають послідовною вибіркою.

Конкретні групи населення можуть мати особливі міркування, які необхідно враховувати при плану-

Блок 1.1 Порівняння проспективних та ретроспективних досліджень

Проспективні когортні дослідження	*Ретроспективні когортні дослідження*
Переваги	**Переваги**
• Збір даних в режимі реального часу • Дозволяє дослідникам шукати відсутні дані, оскільки пацієнт присутній під час збору даних • Дозволяє втручатися і вносити зміни до маршруту пацієнта (у неспостережних дослідженнях)	• Швидше проводити • Менше коштує
Недоліки	**Недоліки**
• Більш тривалий термін • Дорожче • Вимагає більше часу від персоналу	• Важко знайти відсутні дані • Випадки можуть бути пропущені частіше • Для пошуку можуть знадобитися історії хвороби • Дослідники, які оцінюють незалежні/пояснювальні змінні, можуть вже знати про результат

ванні структури дослідження. Прикладом цього може бути старша вікова група, де порушення зору або слуху можуть створювати труднощі з певними методами збору даних, такими як телефонні інтерв'ю.

ВТРУЧАННЯ АБО ВПЛИВ

Втручання є основною змінною в групі лікування. У спостережних дослідженнях пацієнти та лікарі вирішують, яке лікування вони мають отримувати. Оскільки прямого експериментального втручання не відбувається, змінна називається «експозиція» (вплив).

Розглядаючи втручання для хірургічного дослідження, особливу увагу слід приділяти стандартизації між пацієнтами і лікарями. Проведення втручань у стандартизований спосіб має важливе значення для забезпечення порівнянності пацієнтів. Різноманітність способів проведення лікування має бути врахована під час його планування. Наприклад, при дослідженні нової хірургічної техніки або підходу навчання виконанню нового втручання вимагає як часу, так і певних засобів для забезпечення того, щоб усі пацієнти отримували однакове лікування. Стандартизація хірургічного втручання може включати оцінювання в межах дослідження, щоб визначити, чи було досягнуто прийнятного рівня компетентності. Належний протокол дослідження допоможе вирішити це питання (докладніше про протоколи дослідження можна знайти в розділі «Упередженість»).

Прийнятність втручання також повинна бути належним чином обміркована. Якщо втручання є неприйнятним для пацієнтів, буде дуже важко переконати їх (і комітети з питань дослідницької етики) взяти в ньому участь. Одним зі способів забезпечення прийнятності втручань є залучення представників пацієнтів до планування досліджень. Такі представники пацієнтів входять до складу групи розробників дослідження і надають дослідникам зворотний зв'язок щодо найкращих способів забезпечення проведення досліджень у можливий і прийнятний спосіб.

ПОРІВНЯННЯ

Порівняння описує контрольну групу, щодо якої оцінюється втручання або вплив, що становить інтерес. Ця група має бути достатньою мірою порівнянною з групою втручання, щоб можна було зробити достовірні висновки про справжні ефекти будь-якого впливу або втручання.

У контрольованих дослідженнях група порівняння часто використовує *плацебо*, щоб зменшити упередженість і зберегти засліплення. У багатьох дослідженнях група порівняння отримує поточний золотий стандарт клінічної допомоги, щоб створити групу, в якій нове лікування можна порівняти безпосередньо з поточною найкращою або звичайною терапією.

Якщо контрольна група (група порівняння) відрізняється від групи втручання, може виникнути упередженість, що призведе до неправильних висновків. Це викликає особливе занепокоєння в обсерваційних дослідженнях, таких як дослідження «випадок—контроль», де група порівняння обирається дослідниками.

РЕЗУЛЬТАТ

Результат дослідження — це змінна, за допомогою якої вимірюється ефект втручання або впливу, що становить інтерес. Для того щоб результат був корисним, слід враховувати кілька властивостей обраного показника:

- Захворюваність: результат, що нас цікавить, має бути достатньо поширеним. Чим більш поширеним є результат, що нас цікавить, тим менший розмір вибірки необхідний для демонстрації відмінності між групами лікування.
- Безпосередність: результат, що нас цікавить, повинен вимірювати безпосередньо те, чого ми маємо в кінцевому підсумку досягти завдяки втручанню. Наприклад, інтраопераційні зігрівальні пристрої призначені для того, щоб зігріти пацієнта і тим самим зменшити кількість післяопе-

раційних ускладнень. Тому будь-яке дослідження, що вивчає ефективність інтраопераційних зігрівальних пристроїв, повинне використовувати післяопераційні ускладнення як основний результат, а не розглядати різницю в температурі (так званий «сурогатний» результат).

- Визначення: усі результати слід вимірювати за допомогою чітко визначених критеріїв. Бажано, щоб ці критерії широко використовувалися в різних галузях, щоб усі дослідження вимірювали один і той самий результат однаково. Прикладом таких критеріїв можуть бути критерії оцінювання відповіді при солідних пухлинах (RECIST) для визначення прогресування пухлини[6].
- Релевантність: результати мають бути актуальними як для досліджуваних питань, так і для догляду за пацієнтами, хоча те, що може бути важливим для клініцистів, може не бути важливим для пацієнтів. З цією метою при виборі результатів дослідження слід консультуватися з представниками пацієнтів, адже часто якість життя і функціональні розлади мають першочергове значення для пацієнтів.
- Часові межі: результати повинні вимірюватися через відповідні часові інтервали, які відповідають часовим межам, коли очікується отримання результату, що становить інтерес. Наприклад, вимірювання результату через 72 год після процедури може бути релевантним для післяопераційної кровотечі, але не для інфікування місця операції.
- Надійність: обраний показник результату повинен надійно виявляти подію і бути стандартизованим для всіх пацієнтів. Це особливо важливо для багатоцентрових досліджень або широкомасштабних випробувань, де результати оцінюють кілька спостерігачів. Тут наявність стандартизованого засобу оцінювання результату є важливим, оскільки необхідно переконатися, що всі пацієнти оцінюються однаково, інакше результати дослідження можуть бути неточними.

Показники результатів, про які повідомляє пацієнт (англ. *Patient-reported outcome measures*, PROMs), відрізняються від результатів, про які повідомляє дослідник або клініцист. Прикладом PROMs можуть бути валідовані опитувальники загальної якості життя (такі як EQ-5D або HR-QoL). Крім того, PROMs можуть зосереджуватися на конкретних сферах інтересу, таких як сексуальна функція і нетримання сечі, які мають особливе значення в хірургії органів малого таза[7, 8].

Економічна ефективність іноді використовується як показник результативності. Це часто використовується клінічними керівними органами для прийняття рішення про те, чи є лікування оптимальним у плані співвідношення ціни та якості і чи слід його рекомендувати до застосування[9].

Іноді можуть поєднуватися різні результати, утворюючи один комбінований показник. Використання комбінованих показників може дозволити дослідникам виміряти два або більше результатів одночасно, що забезпечує кращу статистичну ефективність для тієї кількості пацієнтів, що беруть участь у дослідженні. Прикладом може бути незапланована госпіталізація до реанімаційного відділення або смерть пацієнта, коли обидва випадки є серйозними ускладненнями, але вони поєднуються для збільшення кількості подій, що зменшує розмір вибірки, необхідний для демонстрації ефективності лікування. Хоча це і корисно, але може призвести до того, що показники результативності будуть складними для інтерпретації.

ДИЗАЙН ДОСЛІДЖЕННЯ

Існує багато схем досліджень, що дають змогу відповісти на клінічні питання. Методологія дизайну дослідження постійно розвивається, проте всі типи досліджень можна умовно розподілити на такі категорії:

- Первинне дослідження (дослідження на рівні окремого пацієнта):
  - Рандомізоване дослідження.
  - Проспективне когортне дослідження.
  - Ретроспективне когортне дослідження.
  - Перехресне дослідження.
  - Дослідження «випадок–контроль».
  - Серія випадків.
  - Звіт про випадок.
- Вторинне дослідження (дослідження, яке враховує кілька джерел первинного дослідження):
  - Систематичний огляд.
  - Систематичний огляд з метааналізом.

Рисунок 1.1, що наведений нижче, надає корисні способи ідентифікації типу дослідження, якщо у звіті нечітко зазначена його структура.

Інший спосіб класифікації доказів – за класифікацією Оксфордського центру доказової медицини (англ. *Centre of Evidence-based Medicine*, CEBM), який поділяє докази відповідно до ризику упередженості. Їх зазвичай називають «рівнем доказовості» (табл. 1.1)[11].

Одним із добре відомих способів класифікації цих досліджень є використання «піраміди доказів», яка забезпечує спрощений спосіб розгляду переваг однієї структури дослідження над іншою. Ця піраміда враховує ті властивості, що є притаманними кожній структурі дослідження, і класифікує їх відповідно до того, наскільки дослідження має дати надійну відповідь на досліджуване питання, максимально наближену до «справжнього значення».

Піраміда доказів супроводжується кількома застереженнями (рис. 1.2). По-перше, не на всі питання можна відповісти за допомогою рандомізованого контрольованого дослідження або через відсутність рівноваги (коли проведення дослідження було б неетичним), або з логістичних причин (коли дослідження було б занадто дорогим або просто нездійсненним). По-друге, погано проведене дослідження може дати неточну або ненадійну відповідь порів-