

СТАТИСТИЧНІ КОНЦЕПЦІЇ

Належна практика наукових клінічних досліджень: деякі джерела хиб

Пилипенко Микола Іванович

Харківський національний медичний університет
МОЗ УкраїниGood clinical trials practice:
some sources of bias

Цель лекции: Познакомить ученых с возможными источниками ошибочных предубеждений в планировании и проведении клинических исследований, которые порождают погрешности результатов.

Материалы и методы: Собственный опыт, полученный благодаря рецензированию рукописей научных статей, рассмотрению проектов диссертационных исследований и клинических научно-исследовательских работ, а также непрерывного ознакомления с текущими публикациями в профессиональных отечественных и мировых журналах, а также обзор ряда публикаций на эту тему.

Вывод: Скрытые источники ошибочных предубеждений могут быть на любом этапе планирования и выполнения исследований, что требует пристального внимания к этой проблеме, чтобы, по возможности, предупредить ошибочные результаты.

Ключевые слова: клиническое исследование, ошибочные предубеждения в исследовании, набор выборки больных, размер выборки, референтный тест.

Мета лекцій: Познайомити науковців з можливими джерелами прихованих хиб упереджень у плануванні і проведенні клінічних досліджень, що призводять до похибок результатів.

Матеріали і методи: Власний досвід, отриманий із рецензування рукописів наукових статей, розгляду проектів дисертаційних досліджень і клінічних науково-дослідних робіт і неперервного ознайомлення з поточними публікаціями у фахових вітчизняних і світових журналах, а також огляд деяких публікацій за цією темою.

Висновок: Приховані джерела хиб упередженості можуть бути на будь-якому етапі планування і виконання дослідження, що потребує пильної уваги до цієї проблеми, щоб, по можливості, запобігти хибним результатам.

Ключові слова: клінічне дослідження, хиби упередженості в дослідженні, набір вибірки хворих, розмір вибірки, референтний тест.

Purpose: To acquaint researchers with possible sources of hidden bias in planning and performing clinical trials that lead to errors in the results.

Materials and Methods: The original experience gained from reviewing manuscripts of scientific papers, dissertation research projects and clinical research and ongoing review of current publications in professional Ukrainian and international journals, as well as review of some relevant publications.

Conclusion: Hidden sources of bias can occur at any stage of the planning and execution of research, which requires careful attention to the problem in order to prevent false results, if possible.

Key words: clinical trial, bias in trial, patients sample recruit, patients sample size, referent test.

Досвід рецензування рукописів наукових статей, розгляду проектів дисертаційних досліджень і клінічних науково-дослідних робіт і, нарешті, неперервного ознайомлення з поточними публікаціями у фахових вітчизняних і світових журналах засвідчив існування потенційних джерел хиб у наукових клінічних дослідженнях, зокрема радіологічних. На мій погляд, для тих з нас, хто виконує наукові дослідження чи планує приступитися до них, важливо бути в курсі цих хиб і знати, як їх уникнути або мінімізувати. В контексті цієї лекції термін *хиба* стосується ситуації, коли отримані з дослідження узагальнені значення вимірів не відповідають таким, що

були б отримані, по-перше, за проведення випробування у *всіх пацієнтів релевантної групи* і, по-друге, при порівнянні результатів вимірювань чи оцінювань, проведених у *репрезентативних вибірках із релевантних груп пацієнтів і адекватно сформованих групах контролю*. За специфікою впливу на підсумкові результати дослідження хиби мають схожість зі звичайними систематичними похибками вимірювань, але в основі свого походження мають упередженість (можливо, мимовільну), яка може бути присутньою на будь-якому етапі проведення дослідження — від складання його дизайну до інтерпретації кінцевих результатів.

Слід підкреслити, що хиби належить відрізняти від вад досліджень, при читанні публікації про результати яких вже на початку можна передбачити тенденційність висновків автора (ів).

На ідею ознайомити вітчизняних лікарів-дослідників з проблемою хиб у дослідженнях наштовхнула вишукана публікація N. A. Obuchowski [1].

Хиби в наборі вибірок пацієнтів

Відразу підкреслюю, що відбір пацієнтів у досліджувані та контрольні вибірки є, безсумнівно, найвідповідальнішою процедурою — виправити результати хиби цього етапу планування дизайну дослідження (чи його здійснення) в подальшому неможливо, і тому дослідження стане марною тратою часу і ресурсів.

Математична статистика пропонує для запобігання упередженості при включенні хворих у дослідну групу застосовувати *принцип випадкового відбору* хворих. Сформована за таким принципом вибірка називається *рандомізованою* (від англ. *at random* — на(в)здогад, навмання, випадково), тобто зробленою (набраною) на здогад, випадково, і, відповідно, дослідження, проведене в такій групі, теж носить назву *рандомізованого*.

Якщо в дизайні дослідження планується порівняти результати вимірювань у досліджуваній та контрольній групах осіб, формування групи контролю рекомендується робити за методом «випадок-контроль», що передбачає знаходити для кожного випадку з досліджуваної групи аналогічний за загальними характеристиками, такими як стать, вік, маса тіла, професія, етнос тощо. Кількість таких характеристик обмежується здоровим глуздом і наявністю контингенту формування вибірки. Головне, що відрізняє контрольну групу від основної, це гарантована відсутність впливу на відібраних осіб фактора, дія якого досліджується в основній вибірці.

Інший (можна сказати, спрощений) варіант формування контрольної групи — рандомізація вибірки, але з урахуванням тих самих характеристик: певна аналогічність за статтю, віком тощо і доведена відсутність дії досліджуваного фактора.

На обґрунтований погляд N. A. Obuchowski [1] одна з хиб при виборі пацієнтів в основну групу дослідження полягає в тому, що можуть відбиратися досліджувані особи селективно за принципом «найхворіші хворі» є «найкращим з хорошо-

го» [2]. Такий відбір хворих зумовлюється упередженням, що при відсутності різниці між такими групами можна буде певно стверджувати про клінічну незначущість досліджуваного діагностичного тесту чи методу лікування. Але це упередження не враховує варіанта результату, коли різниця між групами все ж буде встановлена, то вона буде істотно завищеною, позаяк досліджуваній вибірці відсутні випадки зі «стертими» формами перебігу захворювання, за яких досліджуваний метод (засіб) діагностики чи лікування може мати меншу чи просто *іншу* силу. Наприклад, при оцінці чутливості і специфічності нового діагностичного тесту отримані результати будуть хибними, скоріше за все, завищеними, бо у такому дослідженні враховані далеко не всі варіанти клінічних ситуацій, які стосуються суті дослідження.

Існує поняття *спектр упередженості* [3], під яким розуміється ситуація, коли захворювання, що досліджується, має істотну кількість клінічних форм, і ця обставина ігнорується повністю чи враховується не повною мірою задля бажання скоротити обсяг обстежуваних пацієнтів з різних обставин. Інший варіант спектра упередженості стосується наявності кількох близьких варіантів діагностичного тесту чи методу лікування, а в дослідженні таке диференціювання не визначається з тих же міркувань, що і в попередньому випадку. Цілком очевидно, що спектр упередженості призведе до хибних результатів дослідження, але такій ситуації можна зарадити навіть по завершенні дослідження у запланованому форматі, продовживши дослідження з поступовим формуванням підгруп хворих за ознакою класифікації варіантів захворювання чи тесту.

Рандомізація відбору пацієнтів може бути уявною в ситуації, коли в дослідження включаються пацієнти, що госпіталізуються для лікування в стаціонар лікувального центру, де проводиться дослідження, за направленнями із сторонніх лікувальних закладів. У такому випадку може діяти «невидимий» фактор певної селекції хворих, результатом чого стане нерепрезентативність набраної групи хворих для оцінки захворювання в усій популяції населення. Такий «підводний камінь» можна обійти, маючи певну обізнаність щодо загальних характеристик досліджуваного захворювання і на основі такої поінформованості

виділяючи підгрупи хворих з описом у звіті точних характеристик пацієнтів кожної з них.

Якщо з будь-яких причин виникає необхідність використовувати такі обмежені підгрупи, наприклад, щоб мінімізувати витрати, то необхідно принаймні ретельно збирати і записувати по можливості найдрібніші характеристики кожного пацієнта. Це важливо, перш за все, в тому сенсі, що вимірювання в дослідженні може змінюватися залежно від цих характеристик. Крім того, при повідомленні результатів вимірювань, отриманих у хворих з різними характеристиками, останні обов'язково належить наводити у звіті, щоб інші дослідники мали можливість порівняти характеристики популяції їх пацієнтів з характеристиками вибірки дослідження для узагальнення результатів досліджень для клінічної радіологічної практики.

Найбільш розповсюдженою хвибою є проведення популяційних досліджень з метою з'ясування ефектів радіаційного фактора низького рівня (так званих «малих доз радіації») навколишнього середовища без порівняльної контрольної групи осіб, які проживають в умовах нормального природного радіаційного фону. Ця хиба допускається також при дослідженні захворюваності осіб, професійно пов'язаних з дією радіації в низьких дозах, що не перевищують регламентні рівні опромінювання персоналу. Упередженість таких досліджень полягає в тому, що загальний час проживання на забрудненій території чи стаж роботи в сфері дії радіації низької інтенсивності приймають як час експозиції радіаційного фактора на організм, і всі можливі прояви можливого зростання хворобливості досліджуваного контингенту бездоказово відносять на рахунок зростання накопиченої дози опромінення. Це серйозна омана, оскільки хибні висновки з таких «досліджень» примножують серед населення міфи про виключно негативну дію радіації. В дійсності ж у будь-якій рандомізованій вибірці людей будь-якої професії при пролонгованому динамічному її обстежуванні можна констатувати, щонайменше, ослаблення стану здоров'я, зумовлене таким повсюдним і невідворотним фактором як старінням.

Вибір розміру вибірок

Число осіб для включення в дослідження, тобто обсяг вибірок дослідження, є важливим чинником у дизайні переважної більшості клінічних

досліджень. Обсяг вибірок тісно пов'язаний з потужністю обраної для статистичного аналізу статистики, яка здатна виявити статистично значущу різницю між вибірками, якщо така дійсно є. Існує компроміс між можливим розміром вибірки та адекватною статистичною потужністю. Стратегії скорочення обсягу вибірки при збереженні розумної потужності є важливим елементом дизайну дослідження. На щастя, багато досліджень в радіології мають простий дизайн, для якого визначення належного обсягу вибірки відносно просте. Крім того, існують численні веб-сторінки [4–6] і комп'ютерні програми [7–9], розроблені для розрахунку необхідного розміру вибірки при використанні для аналізу параметричних і непараметричних статистик.

Хиби у виборі референтного методу лікування чи діагностичного тесту

Вочевидь, надзвичайно важливим у порівняльних дослідженнях є вибір методу чи засобу, з ефективністю якого проводять порівняння досліджуваного. Метод чи засіб, з яким проводиться порівняння нового, називається *референтним*, або *опорним*. Наразі для більшості клінічних ситуацій розроблені протоколи чи стандарти діагностики і лікування захворювань, деякі з котрих отримали статус «золотого стандарту». Цілком зрозуміло, що в якості референтного методу діагностики чи засобу лікування в першу чергу має бути розглянуто застосування того з них, який входить до рекомендованого протоколу ведення хворого в клінічній ситуації, аналогічній передбачуваній дизайном дослідження. Але, в усякому разі, діагноз і стан хворого, який планується до включення в дослідження, мають бути в будь-який спосіб верифіковані майже на 100%, незалежно від того, увійде він у досліджувану вибірку чи вибірку порівняння. Це є одним з базових положень щодо якості організації і проведення клінічних досліджень.

Відома практика порівняння ефективності нового лікарського засобу з плацебо. Навряд чи можна назвати етичним такий вибір порівняння, оскільки інформація, отримана в такому дослідженні, матиме досить мізерне клінічне значення (встановлення лише факту дієвості чи недієвості досліджуваного лікарського засобу в певних клінічних ситуація, за яких ефективна медична допомога вірогідно могла б бути надана існуючим

і доступним засобом) і ціною їй буде фактична відмова групі хворих у лікуванні. Виправданням цьому не може бути і той факт, що плацебо буває ефективним в середньому у 10 % і більше випадків, що пояснюється ефектом навіювання. Крім того, застосування плацебо як «лікарського засобу порівняння» створює етичну колізію. Відповідно до кодексу «Належної клінічної практики» хворий має бути поінформованим дослідниками щодо усіх деталей дослідження, отже і про застосування в дослідженні плацебо. Але в такому разі пацієнт навряд чи погодиться на участь в такому дослідженні, і дослідник не має іншого вибору, як у якийсь спосіб приховати цей факт. В усякому разі така практика проведення дослідження може бути прийнятною тільки в дуже обмеженій кількості специфічних випадків (наприклад, коли дієвих засобів лікування хвороби не існує), які потребуватимуть всебічного серйозного обґрунтування.

Якщо проводити порівняльне вимірювання нового діагностичного тесту (наприклад, вимірювання його чутливості і специфічності або порівняння його точності), то для порівняння належить вибрати «золотий стандарт». Але хибою упередження в такому випадку може стати неврахування тієї обставини, що цей референтний тест не може, як те належить, забезпечити 100% верифікацію стану всіх хворих, що відбираються в дослідження. Отже, вибір опорної стандартної процедури верифікації стає найскладнішою частиною проектування дослідження. Дехто стверджує, що не існує такого поняття, як «золотий стандарт» в значенні ідеального тесту. Навіть результати патологоанатомічного аналізу не на 100% достовірні, тому що як і радіологія, патанатомія є тлумачною дисципліною [10]. Саме тому в розв'язанні такої проблеми потрібна дециція здорового глузду.

Іноді в такій ситуації можна скористатися літературними даними щодо встановлених значень специфічності і чутливості опорного діагностичного тесту. Але має бути упевненість в їх достовірності. Найбільш продуктивний шлях (та не легший) — верифікація на основі сумісного використання анамнезу, фізикального обстеження хворого, інструментальних й лабораторних даних, зокрема біопсійної патоморфології, перебігу захворювання, результатів лікування, подальшого нагляду тощо.

Хиби у проведенні та інтерпретації результатів дослідження

У буденній практиці зазвичай радіологічні дослідження хворого інтерпретуються і навіть інколи виконуються з урахуванням даних, уже отриманих з інших радіологічних і не радіологічних тестів. Наприклад, рентгенолог робить знімки хворому та інтерпретує їх, знаючи результати ультразвукового чи радіонуклідного дослідження цього хворого. Це звичайна практика. Але при проведенні дослідження на оцінку точності тесту такого не повинно бути, інакше попередні відомості створюють інтерпретаційну упередженість і, як наслідок, систематичну похибку вимірювання. Щоб уникнути такої хиби, дослідження має виконуватися у «сліпий» спосіб, тобто без попередньої поінформованості.

Упередженість такого типу часто виникає при ретроспективних дослідженнях, за яких дослідник отримує увесь готовий пакет діагностичних зображень, виконаних у різні способи. Ще більш значна упередженість виникатиме, якщо в розпорядженні дослідника буде два чи більше пакетів зображень, виконаних через істотні проміжки часу.

Вочевидь, єдиний дієвий спосіб запобігти такій упередженості це надати можливість читання та інтерпретації (з невідкладним написанням звіту) у першу чергу зображень, отриманих оцінюваним способом, групі фахівців.

Ще одне джерело упередженості виявили Egglin T.K. а. Feinstein A.R. [11]. Вони встановили, що «чутливість читача» зображень істотно збільшується, якщо в серії розглядуваних зображень, отриманих у різних хворих з різною патологією за допомогою досліджуваного методу чи способу, частота зображень з досліджуваною патологією збільшується. Як висновок, вони пропонують у дослідженні створювати серію знімків, серед яких зображення досліджуваного захворювання відповідатиме частоті цього захворювання в загальній популяції хворих.

Отже, з викладеного вище робимо висновок, що приховані джерела хиб упередженості можуть бути будь-де, що потребує пильної уваги до цієї проблеми при плануванні клінічного дослідження. Дослідник має докласти всіх зусиль, щоб уникнути хиб упередженості, або ж мінімізувати їх вплив на результати дослідження.

Цитування

1. *Obuchowski N. A. Special Topics III: Bias.* // *Radiology*, 2003; vol. 229: P. 617–621.
2. *Sox H., Stern S., Owens D., Abrams H.L. Assessment of diagnostic technology in health care: rationale, methods, problems, and directions.* Washington, DC: National Academy Press, 1989.
3. *Ransohoff D.F., Feinstein A.R. Problems of spectrum and bias in evaluating the efficacy of diagnostic tests.* // *N Engl J Med* 1978; 299:926–930.
4. *Bond J. Power calculator.* Available at: <http://calculators.stat.ucla.edu/powercalc/> Accessed March 11, 2003.
5. *Uitenbroek DG. Sample size: SISA—simple interactive statistical analysis.* Available at: <http://home.clara.net/sisa/samsize.htm>. Accessed March 3, 2003.
6. *Lenth R. Java applets for power and sample size.* Available at: www.stat.uiowa.edu/~rlenth/Power/index.html. Accessed March 3, 2003.
7. *NCSS Statistical Software. PASS 2002.* Available at: www.ncss.com/pass.html. Accessed March 3, 2003.
8. *SPSS. Sample Power.* Available at: www.spss.com/SPSSBI/SamplePower/. Accessed March 3, 2003.
9. *Statistical Solutions. nQuery Advisor.* Available at: www.statsolusa.com/nquery/nquery.htm. Accessed March 3, 2003.
10. *Valenstein P.N. Evaluating diagnostic tests with imperfect standards.* // *Am J Clin Pathol* 1990; 93:252–258.
11. *Eggin TK, Feinstein AR. Context bias: a problem in diagnostic radiology* // *JAMA* 1996; 276:1752–1755.

Надходження до редакції 21.10.2013.

Прийнято 21.10.2013.

Адреса для листування:

Пилипенко Микола Іванович,
ДУ «Інститут медичної радіології ім. С.П. Григор'єва
НАМН України»,
вул. Пушкінська, 82, Харків, 61024, Україна