

М.А.Іванчук, П.Р.Іванчук, В.К.Тащук

ВИКОРИСТАННЯ СТАТИСТИЧНИХ МЕТОДІВ В МЕДИКО-БІОЛОГІЧНИХ ДОСЛІДЖЕННЯХ

Кафедра кардіології та функціональної діагностики (зав. – проф. В.К.Тащук),
кафедра медичної та біологічної фізики (зав. – проф. М.В.Шаплавський)
Буковинського державного медичного університету, м. Чернівці

Резюме. У роботі описані основні помилки, які трапляються при використанні статистичних методів у клінічних та експериментальних дослідженнях. Наведено поради в оцінці результатів медико-біологічних досліджень.

Ключові слова: статистичні методи, рівень значущості, критерій Ст'юдента, параметричні критерії, непараметричні критерії, нормальний закон розподілу.

Статистичні методи в клінічних дослідженнях стали використовуватися відносно недавно – з 30-х років ХХ сторіччя. З того часу кількісно-статистичний підхід до опису й аналізу клінічних явищ міцно затвердився у свідомості багатьох поколінь медиків. Саме біостатистика зіграла ключову роль у переході медичних досліджень від опису окремих спостережень і серій випадків до проведення експериментальних робіт з використанням контрольних груп і масштабних рандомізованих контрольованих випробувань, що стало науковим стандартом якості наукових досліджень. Формулювання основного питання дослідження, вибір відповідного питання методу і способу організації дослідження, особливості відбору

хворих і характер отриманих даних – все це визначає вибір адекватного методу статистичного аналізу і впливає на достовірність отриманих даних [1].

При цьому одні автори намагаються свідомо застосовувати статистичні методи, щоб на основі результатів аналізу вибіркового спостереження зробити висновок відносно всієї популяції, що досліджується. Інші використовують статистичний аналіз тому, що так робить більшість їх колег, або вважаючи, що наявність статистичних висновків піднімає престижність публікації, викликає більше довіри до неї. Треті – задовольняючи вимогам одного із співавторів, як правило, наукового керівника основного автора. Зазвичай

такий керівник, ще на початку своєї наукової кар'єри, вдало застосував той чи інший метод аналізу, і з тих пір намагається відтворювати його в наступних роботах [5].

Нами було проаналізовано 107 статей із журналу „Буковинський медичний вісник” з розділів клінічних та експериментальних досліджень (з них 59 статей за 1998 рік та 48 за 2003 рік), в яких використовувалася статистична обробка матеріалу. Аналіз засвідчив відтворення однакових помилок математичного обґрунтування наукових досліджень.

У 50 (47%) статтях автори взагалі не вказували статистичні методи, які використовували. При цьому в тексті наявні вирази типу « $p < 0,05$ » та « $p < 0,01$ ». Тоді як у контексті задач, що описуються, вирази « $p < 0,05$ » та « $p < 0,01$ » можуть бути отримані лише при використанні конкретних статистичних критеріїв, наприклад, критерію Пірсона χ^2 , критерію Уїлкоксона і т.д. [4].

У 30 роботах трапляються загальні фрази, що, мабуть, на думку авторів статей, описують використані статистичні методи: „Аналіз отриманих результатів проводився методом варіаційної статистики з використанням персонального комп'ютера”; „Всі результати підлягали комп'ютерній статистичній обробці”; „Загальногрупові характеристики отримано з використанням статистичних програм”; „Отримані результати досліджень обробляли за допомогою методів варіаційної статистики на мікро-ЕВМ”; „Отримані результати оброблені статистично за спеціальними програмами”; „Цифровий матеріал обробляли з використанням стандартних статистичних комп'ютерних програм”; „Методом клініко-статистичних досліджень вивчено ...”; „Математичний розрахунок показників проводили за відсотковими даними без побудови графіка”. Для математика ці фрази звучать так само, як для медика фрази „Хворого лікували за допомогою хірургічних методів скальпелем” або „Ліки призначили з використанням стандартних рецептурних довідників”. Погодьтеся, якщо в списку використаних методів не згадується деякий конкретний метод і мета його використання, то немає змісту обговорювати отриманий за його допомогою результат, і навпаки [5].

У 20 статтях вказано комп'ютерні програми, за допомогою яких проводилися обчислення, або просто тип комп'ютера. Найчастіше автори посилаються на програми „Statgraphics”, „Microstat” „BioStat” „Statistica”, „Excel”, нерідко з помилками: *Exell*, *Exel*, *Excell*, *Statgrafics*, *Statgraphycs*. Наведемо приклади описання статистичних методів, які використовувалися: „Математичні розрахунки та статистичний аналіз проводили на IBM AT-286 з використанням пакета програм „Statgraphics”; „Отримані результати оброблені статистично за програмою „Microstat” (США)”; „Статистичний аналіз даних проводили на комп'ютері IBM PC AT 386 DX за допомогою програми „Statgrafics”; „Статистичну

обробку даних, включаючи кореляційний і регресійний аналіз проводили на комп'ютері IBM PC AT 386 DX за допомогою програми Statgrafics”; „Статистичну обробку проводили за допомогою програми MS® Excel® 2000™ (9.0.2812)”; „Статистична обробка отриманих даних проведена на PC IBM 486 за допомогою баз даних Excell-5 і програми „Statgraphics” (США)”; „Статистична обробка результатів проводилася на комп'ютері за допомогою пакета програм статистичної обробки Microsoft Excel 5.0.”. Відмітимо, що будь-який пакет статистичних програм містить не один десяток методів обробки інформації, кожен з яких можна використовувати за певних умов, і кожен з яких дасть відмінний від інших результат статистичної обробки матеріалу. Крім того, програму Excel не коректно називати базами даних або програмою статистичної обробки. Excel – це електронні таблиці, які дозволяють обробляти і зберігати будь-яку інформацію, здійснювати складні математичні обчислення та будувати графіки. Статистична обробка інформації – це одна з багатьох можливостей цієї програми, на відміну від суто статистичних програм „Statgraphics”, „Microstat”, „Biostat”, „Statistica”. Зауважимо, що результат обробки даних (не тільки статистичної, але будь-якої) не залежить від типу та потужності обчислювальної машини, на якій вона проводилася. Інформація про тип ЕОМ є зайвою, і немає змісту приводити її в дисертації чи в статті [4].

До речі, у роботі, де „Статистична обробка результатів проводилася на комп'ютері за допомогою пакета програм статистичної обробки Microsoft Excel 5.0”, ми не виявили жодного статистичного показника, жодного числового опису матеріалу та результату статистичної обробки матеріалу.

Роботи, в яких метод був описаний, не відрізняються оригінальністю – в 41 (72%) з них був використаний критерій Стьюдента. Відмітимо, що використання цього критерію вимагає виконання двох умов - вибірки нормально розподілені, а їх генеральні дисперсії рівні між собою. Досвід показав, що при клінічних дослідженнях вибірки дуже рідко бувають розподілені нормально, а центральна гранична теорема, в якій стверджується, що при великому об'ємі вибірки її розподіл прямує до нормального, справедлива для дійсно великих вибірок (об'ємом більше 100 елементів), а в роботах, які нами аналізувалися, групи переважно склалися з 10-40 осіб, а іноді з 5-7. Крім того, перед використанням критерію Стьюдента необхідно за допомогою критерію Фішера перевірити рівність дисперсій. Посилання на використання критерію Фішера ми побачили лише у 3 статтях, і в жодній статті не було вказано, чи перевірялися розподіли вибірок на нормальність.

Критерій Стьюдента, який дуже легко обчислюється навіть вручну, став своєрідною панацеєю для медиків. Складається враження, що авто-

ри статей просто не знають про існування великої кількості інших критеріїв і методів і вважають критерій Стьюдента загальноприйнятим методом: „*Статистична обробка результатів досліджень здійснювалась загально визначеним методом варіаційної статистики із застосуванням критерію Стьюдента*”; „*Результати піддавали загальноприйнятій статистичній обробці, вірогідність відмінностей визначали за t-критерієм Стьюдента*”.

Які ж ще методи використовувалися в роботах? У 10 статтях проводився кореляційний аналіз (в 7 з них будувалася пряма регресія). У 6 роботах використовувався критерій Уїлкоксона-Манна-Уїтні, у 8 статтях використовувалися інші методи – χ^2 -критерій (кримські автори), метод найменших квадратів, дисперсійний аналіз, метод Краскелла-Уолліса (запорізькі автори), спектральний аналіз, метод Коена, Консинор-аналіз. Тобто, із 107 статей критерій Стьюдента не використовували лише в 24 (22%).

Дуже сумно за працю великих учених, які чотири сторіччя розвивали математичну статистику, створювали нові теорії, формулювали та доводили теореми, розробляли нові методи та критерії. І все для чого? Щоб їх нащадки користувалися лише одним критерієм, розробленим англійським хіміком Уільямом Госсетом, який працював на пивоварному заводі та публікувався під псевдонімом "Student", та доведеним до медичного суспільства після публікації в 1960 році в роботі Н.А.Ойвін [6].

Можна без перебільшення сказати, що в даний час бездумне застосування t-критерію Стьюдента в більшості біомедичних публікацій приносить більше шкоди, ніж користі [4].

Не можна стверджувати, що в усіх роботах опис використаних статистичних методів був некоректним. Наведемо найбільш математично обґрунтований опис, який ми прочитали в роботі запорізьких науковців: „Дані представлені у вигляді: вибіркове середнє значення \pm стандартна помилка середнього значення. При перевірці статистичних гіпотез нульову гіпотезу відкидали при рівні значущості менше 0,05. Наявність і вірогідність відмінностей між вибіркковими середніми величинами оцінювали за допомогою непараметричного критерію Kruskal-Wallis з подальшим попарним порівнянням груп за допомогою критерію Dunn. Для оцінки ступеня взаємозв'язку між парами незалежних ознак використовували коефіцієнт рангової кореляції R.Spearman. Для вивчення наявності та характеру залежності між ознаками, що вивчаються, використовували процедуру множинного покрокового регресійного аналізу з виключенням лінійної моделі та покроковим виключенням з моделі статистично незначущих факторів”.

Багато казусів трапляється при описанні величини p . Так, в одній із робіт автори вказують, що „ p - ступінь вірогідності різниць показників, що вивчаються, в порівнянні з контролем”. Віро-

гідність, як відомо, може приймати значення від 0 до 1. Тобто, виходячи з висловлення авторів статті, чим ближче p до 1, тим більше відрізняються показники від контролю, і чим ближче значення p до 0, тим менше показники відрізняються від контрольних. Але нижче автори кажуть, що вважають вірогідними зміни показників при $p < 0,001$. В іншій статті автори стверджують, що „ p - достовірність отриманих змін” і при цьому вважають зміни достовірними при $p < 0,01$. Тобто, автори дозволяють собі помилитися в 99% випадків. У роботі, де автори шукали кореляційний зв'язок між змінними, було сказано, що „вірогідність кореляційного зв'язку $p < 0,001$ ”, тобто, з імовірністю менше 0,1% можна стверджувати, що коефіцієнт кореляції був знайдений правильно.

У 88 зі 107 статей автори використовували величину p , і в 50 з них навіть не вказали, що таке p . У 32 роботах величина p була описана некоректно: „ступінь вірогідності різниць показників, що вивчаються в порівнянні з контролем”, „ступінь вірогідності змін між досліджуваними групами”, „вірогідна різниця порівняно з показниками на початку лікування”, „вірогідність різниці по відношенню до $M_{\pm m}$ ”, „достовірність отриманих змін”, „вірогідність кореляційного зв'язку”, і лише в 5 роботах не було особливих суперечностей між використанням та описанням величини p : „вірогідність порівняно з контрольною групою”. Знову порадували нас запорізькі колеги, вказавши, що „ p – рівень значущості”.

Вищенаведене дозволяє обґрунтувати наступні поради в оцінці результатів проведених досліджень.

1. Зазвичай у наукових роботах необхідно довести відмінність між двома групами об'єктів дослідження. Наприклад, є група пацієнтів, що отримували певне лікування та група пацієнтів, що не приймали такого лікування. При цьому дослідник спочатку висуває гіпотезу про те, що групи не відрізняються. Якщо імовірність того, що ці групи однакові, менше 5%, дослідник сміливо може відкинути гіпотезу про подібність груп і стверджувати, що групи вірогідно відрізняються. Цей п'ятивідсотковий поріг називається рівнем значущості. Як відомо, в медико-біологічних дослідженнях рівень значущості зазвичай приймають рівним 0,05; 0,01; 0,001 або 5%; 1%; 0,1% відповідно.

Отже, p – це рівень значущості або ймовірність того, що досліджувані групи не відрізняються, або ймовірність подібності основної групи та групи контролю. А ступінь вірогідності різниць показників або ступінь вірогідності змін між досліджуваними групами – це величина, що дорівнює $1-p$. Тобто, при $p=0,05$ ступінь вірогідності різниць показників дорівнює 0,95.

Крім того, у виразах типу „показники достовірно відрізняються”, „показник достовірно зростає” замість слова „достовірно” краще викорис-

товувати словосполучення “статистично значимо”: “показники статистично значимо відрізняються”, “показник статистично значимо зростає”.

2. При виборі статистичного критерію, який дослідник буде використовувати, у першу чергу розподіли всіх вибірок необхідно перевірити на нормальність. Якщо в пакеті статистичних програм, які використовуються, такої перевірки немає, її легко здійснити за допомогою звичайного Excel.

Спочатку необхідно обчислити середнє значення M (відповідна функція в Excel називається СРЗНАЧ), медіану (МЕДИАНА) и моду (МОДА) вибірки. При нормальному розподілі ці три величини мають бути рівними між собою. Потім необхідно знайти коефіцієнт асиметрії (функція SKOS) та коефіцієнт ексцесу (ЭКСЦЕСС). Ці величини мають бути близькими до нуля. Якщо ці умови істотно не виконуються, розподіл вибірки не можна вважати нормальним.

Якщо ж умови виконуються, необхідно перевірити ще одну умову (так зване „Правило трьох сигм”). Необхідно знайти середнє квадратичне відхилення вибірки s (функція СТАНДОТКЛОН). При нормальному законі розподілу:

- майже всі значення вибірки попадають в інтервал $(M-3 \times s; M+3 \times s)$;
- дві третини значень вибірки попадають в інтервал $(M-s; M+s)$;
- половина значень вибірки попадають в інтервал $(M-0,625 \times s; M+0,625 \times s)$.

Якщо ця умова виконується, розподіл вибірки можна вважати нормальним, якщо умова не виконується, розподіл вибірки не є нормальним.

3. Якщо розподіл вибірки нормальний, використовують параметричні критерії („славнозвісний” критерій Стьюдента-Фішера), якщо ненормальний, необхідно використовувати непараметричні критерії (критерій знаків, Уїлкоксона, Ван дер Вардена, Колмогорова-Смірнова, Фішера, Розенбаума).

4. Перед використанням критерію Стьюдента необхідно за допомогою критерію Фішера перевірити гіпотезу про рівність генеральних дисперсій двох вибірок. Зокрема, в Excel це можна зробити, викликавши функцію ФТЕСТ. Результатом обчислення цієї функції буде ймовірність того, що дисперсії вибірок не відрізняються. І лише в тому випадку, коли обидві вибірки розподілені за нормальним законом, а їх дисперсії статистично вірогідно не відрізняються, можна використовувати критерій Стьюдента.

5. Якщо хоча б одна з вибірок розподілена ненормально, критерій Стьюдента не можна використовувати. Ви можете вибрати будь-який із непараметричних критеріїв, які входять у пакет статистичних програм, якими Ви користуєтесь. Ми наведемо приклад використання непараметричних критеріїв Уїлкоксона.

6. Якщо вибірки незалежні між собою (група хворих та група контролю, різні вікові групи, група чоловіків та група жінок), використовують

U-критерій Уїлкоксона (його ще називають критерієм Уїлкоксона-Манна-Уїтні). Спочатку дві вибірки об’єднують в одну і знаходять ранги (або, іншими словами, порядкові номери) елементів об’єднаної вибірки. Після цього знаходять суми рангів елементів окремо першої та другої вибірки. Суму рангів елементів першої вибірки позначають R_1 , другої – R_2 . Після цього знаходять величини

$$U_1 = n_1 n_2 + \frac{n_2(n_2 + 1)}{2} - R_2$$

$$U_2 = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - R_1$$

та

де n_1, n_2 – об’єми вибірок. Найбільшу з цих величин порівнюють з критичним значенням. Критичні значення знаходять за спеціальними таблицями [3] або, якщо об’єм вибірки більше 20, обчислюють за формулами

$$U_{0,05} = 0,45 n_1 n_2 - 1,18(n_1 + n_2) + 6,45$$

для рівня значущості $p=0,05$ та

$$U_{0,01} = 0,43 n_1 n_2 - 1,74(n_1 + n_2) + 11,13$$

для рівня значущості $p=0,01$. Якщо найбільша з величин U_1 та U_2 більша за критичне значення, то вибірки відрізняються з відповідним рівнем значущості.

7. У випадку залежних вибірок (хворі до та після лікування) використовують T-критерій Уїлкоксона. Як критерій використовується найменша з двох величин – R_1 (сума рангів додатних різниць між членами вибірок) та R_2 (сума рангів від’ємних різниць між членами вибірок). Критичне значення знаходять за спеціальними таблицями [3] або, якщо об’єм вибірки більше 25, обчислюють за формулою

$$T = \frac{n(n+1)}{4} - t_\alpha \sqrt{\frac{n(n+1)(2n+1)}{24}}$$

де t_α значення t -критерію Стьюдента для відповідного рівня значущості α . Якщо найменша з величин R_1 та R_2 менша за критичне значення, то вибірки відрізняються з відповідним рівнем значущості.

Для отримання більш детальної інформації про використання статистичних методів у клінічних та експериментальних дослідженнях можна звернутися до класичної роботи з розрахунків у медичних дослідженнях [3] та до оригінальної роботи американського автора, в якій просто, зрозуміло для нефакхівців мовою пояснюються основи математичної статистики та принципи їх використання в медицині [2].

Література

1. Бацинский С.Е. Статистика умеет много гитк //Междунар. ж. мед. практики. – 1998. - №4. - С.13-15.

2. Гланц С. Медико-биологическая статистика. – М.: Практика, 1999. – 460 с.
3. Лакин Г.Ф. Биометрия: Учебное пособие для биологов спец. вузов. – 4-е изд., перераб. и доп. – М.: Высш. школа, 1990. – 352 с.
4. Леонов В.П. Долгое прощание с лысенковщиной // <http://www.biometrica.tomsk.ru/lis.htm>
5. Леонов В.П. Три "Почему ..." и пять принципов описания статистики биомедицинских публикациях // <http://www.biometrica.tomsk.ru/principals.htm>
6. Ойвин И.А. Статистическая обработка результатов экспериментальных исследований // Патол. физиол. и эксперим. терапия. – 1960. – Т.4, №4. – С.76-85.

USING STATISTICAL METHODS IN MEDICAL AND BIOLOGICAL RESEARCHES

M.A.Ivanchuk, P.R.Ivanchuk, V.K.Tashchuk

Abstract. The paper describes the basic mistakes that occurred when using statistical methods in clinical and experimental studies. Pieces of advice have been given pertaining an evaluation of the findings of medico-biological studies.

Keywords: statistical methods, significance level, Student's criterion, parametric criteria, nonparametric criteria, normal distribution law.

Bukovinian State Medical University (Chernivtsi)

Buk. Med. Herald. – 2005. – Vol.9, №4.- P.123-127

Надійшла до редакції 13.04.2005 року