

Шановні читачі! У рубриці „Методологія наукових досліджень” редакція публікує матеріали, що пов’язані з найважливішими аспектами наукової діяльності: організаційно-методичним забезпеченням наукових видань, загальним принципам статистичного, біометричного і математичного супроводження досліджень, а також оригінальними методичними підходами вітчизняних і зарубіжних морфологів. У цьому випуску редакція продовжує розглядати питання, які пов’язані з методологією підготовки матеріалів для публікації результатів дослідження у наукових журналах.

ФОРМАТ СУЧАСНОЇ ЖУРНАЛЬНОЇ ПУБЛІКАЦІЇ ЗА РЕЗУЛЬТАТАМИ КЛІНІЧНОГО ДОСЛІДЖЕННЯ.

ЧАСТИНА 4. БІОСТАТИСТИКА

К.П.Воробйов

(Луганська обласна клінічна лікарня)

Матеріали перекладені українською і публікуються за джерелом:

Воробьев К. П. Формат современной журнальной публикации по результатам клинического исследования. Часть 4. Биostatистика / К. П. Воробьев // Украинский медицинский часопис. - 2008. - № 6. - С. 79-91.- http://www.umj.com.ua/archive/68/pdf/1109_rus.pdf

Вступ

У попередніх трьох частинах цієї серії публікацій позначено сутність проблемної ситуації (Воробйов К.П., 2007), коротко описані міжнародні рекомендації по підготовці біомедичних журнальних публікацій (ЖП) (Воробйов К.П., 2008а) і представлені загальні дані про дизайн клінічного дослідження (КД) як найбільш важливу і порівняно нову методологічну проблему в клінічній науці (Воробйов К.П., 2008б). При плануванні КД дослідник повинен передбачити і мінімізувати двох типів помилок: систематичні і випадкові. Якісний дизайн КД забезпечує зниження вірогідності появи систематичних помилок, а мінімізація випадкових помилок забезпечується вживанням методів статистичного аналізу, які адекватні клінічному матеріалу і цілям дослідження. У міру використання обчислень в КД методи статистичного аналізу придбали певну специфічність і тому зазвичай сукупність цих методів називають біостатистикою.

У останні десятиліття методи біостатистики активно обговорюються у всьому світі. Це пов’язано з підвищенням складності вживаних обчислювальних підходів в КД і з появою нових програмних засобів для реалізації обчислень. Як дослідникові, так і лікаріві потрібно все більше спеціалізованих знань біостатистики для якісного представлення результатів КД і для розуміння нової епідеміологічної мови вірогідності клінічних явищ при читанні публікацій. У країнах пострадянського простору склалися особливі традиції використання статистики в біомедичних дослідженнях, історія яких цікаво описана в статті професійного статистика В.П.Леонова «Долгое

прощание с лысенковщиной» (<http://www.biometrica.tomsk.ru/lis.htm>). Тоді як в розвинених країнах формувалися і розвивалися традиції якісного використання біостатистики, в країнах пострадянського простору в середовищі клініцистів сформувався зверхнє відношення до ролі обчислювальних методів в медицині. У Росії останніми роками з’явилася критичне число фахівців у біостатистиці, що дозволяє колективно обговорювати проблему. Досвід російських колег корисний і повчальний для нашої країни, оскільки в Україні ці питання фактично не обговорюються.

Мета дослідження — провести системний аналіз проблем якісного використання методів біостатистики в клінічній ЖП і представити найбільш грубі типові помилки статистичного аналізу в КД.

СУТНІСТЬ ПРОБЛЕМНОЇ СИТУАЦІЇ

Тема якості представлення статистичного матеріалу в ЖП є усе більш актуальною в клінічному науковому співтоваристві з кількох причин. З одного боку, відбувається лавиноподібне збільшення потоків інформації в медичній галузі, з іншого — в останні два десятиліття істотно змінився погляд на теорію КД у зв’язку з розвитком епідеміологічних методів при вивченні захворювань (Флетчер Р. і співавт., 1998). У свою чергу сучасні епідеміологічні методи засновані виключно на аналізі числової інформації. Паралельно з розвитком методології КД відбувається ще стрімкіший розвиток інформаційних технологій, а значить, і обчислювальних можливостей сучасних інструментів статистичного аналізу. В результаті таких стрімких змін, при зростаючій спеціалізації наук і домінуванні так званої мозаї-

чної освіти в сучасному науковому співтоваристві дослідник частенько не в змозі зрозуміти і освоїти нові інструменти обчислювального аналізу. Окрім цього, інертність людських звичок на тлі зростаючої геронтологізації правлячої наукової еліти в країнах пострадянського простору, за відсутності зворотних зв'язків в системі оцінки рейтингу вченого — все це створює край несприятливі умови для розвитку методології КД.

Представимо аналіз загальних проблем якісного використання біостатистики в ЖП на основі російських досліджень, оскільки в Україні ця тема навіть не обговорюється. Цікавий матеріал по даній темі опублікований в «Міжнародному журналі медичної практики» (№ 2, 2006). Це видання представляє аналіз проблемної ситуації в області біостатистики з точки зору провідних фахівців Росії. Потім в одному з наступних номерів журналу (№ 1, 2007) публічно обговорюється відкритий лист одного з російських фахівців з біостатистики.

У статті С.Л.Плавінського (Санкт-Петербурзька медична академія післядипломної освіти) на основі аналізу вживання біостатистики в ЖП висміюється відношення переважної частини лікарів-дослідників до використання різних статистичних методів. Це відношення частенько нагадує релігійні обряди, в яких **шаблонний підхід і віра в непогрішимість незрозумілих для дослідника знаків і символів** замінює розуміння основ статистики і усвідомлений вибір статистичних методів відповідно до завдань дослідження. Автор вважає невірним повністю передоручати цю важливу частину методології КД спеціалістам-статистикам (Плавінський С.Л., 2006).

У наступній публікації А.Н. Вараксін з Лабораторії промислового моделювання Інституту промислової екології Уральського відділення РАН на основі досвіду взаємодії з дослідниками-медиками констатує, що практично **ніхто з медиків не розуміє сенс використовуваних статистичних методів і вихідних параметрів**, повсюдно використовуються певні шаблони невірних підходів до аналізу даних і статистичних формулювань. Автор вважає, що необхідна спільна робота лікаря-дослідника і фахівця із статистики. Суть відношення автора до проблеми полягає в тому, що «... ми не заперечуємо, що біологи і медики в деяких випадках можуть побудувати правильну математичну модель. Істина ж полягає в тому, що сам біолог або медик не знають, яка з побудованих ними моделей є адекватною» (Вараксін А.Н., 2006). Тому, на думку автора, якщо на початку дослідження співпраця із статистиком є бажаною, то в кінці — обов'язковим.

Наступний експерт з Санкт-Петербурзької медичної академії післядипломної освіти на основі аналізу результатів спілкування з аспіран-

тами виділяє три групи молодих учених, кожна з яких передбачає наступні підходи у вирішенні майбутніх проблем із статистичною обробкою їх дисертаційного матеріалу. Найбільш багаточисельна група покладає надійні на своїх маститих керівників, які мають «досить добре апробовані шаблони такого оформлення», друга група орієнтована на фахівців, які надають послуги з обробки результатів, третя група — це ті, хто вже добре володіє інформаційними технологіями, але переоцінюють свої можливості, покладаючи всі надії на сучасні статистичні програми і нехтуючи теорією статистики (Дюк В.А., 2006).

У наступній публікації три викладачі біостатистики з різних установ Санкт-Петербурга заявляють, що «... у вітчизняних медичних дослідженнях дуже часто спостерігається або ігнорування сучасних можливостей статистичних методів або їх безграмотне використання» (Барт А.Г. і співавт., 2006). Вони також акцентують увагу, що і практикуючий лікар повинен розуміти сучасну статистичну мову і знати методологію сучасної медицини. У іншому випадку лікар не зможе ефективно поповнювати свої знання при читанні сучасних ЖП. Автори цієї публікації представили приклад того, яким чином поліпшити положення справ у використанні біостатистики: ними був проведений формальний аналіз якості використання біостатистики в авторефератах дисертацій, що послужило підставою для введення в Санкт-Петербурзькому державному медичному університеті ім. І.П. Павлова обов'язкового теоретичного курсу основ біомедичної статистики. У цьому питанні автори публікації посилаються на досвід розвинених країн: практично у всіх медичних університетах Європи і Америки є або кафедри, або відділи біостатистики. Зазвичай ці підрозділи знаходяться у складі кафедри епідеміології, співробітники яких читають для студентів і аспірантів медичних спеціальностей багаторівневі курси, а також ведуть підготовку за фахом «біостатистика». У резюме даної публікації констатується, що «В даний час рівень статистичної освіти російських медиків край низький» (Барт А.Г. і співавт., 2006).

Серед цих публікацій, на наш погляд, на особливу увагу заслуговує думка відомого в Росії фахівця з біостатистики, головного редактора сайту «Біометрики» (<http://www.biometrika.tomsk.ru>), кандидата технічних наук, доцента державного університету Томська В.П. Леонова. Вже багато років Василь Петрович веде активну аналітичну і експертну роботу по вивченню стану використання методів біостатистики в біомедичних дослідженнях, проводить регулярні виїзні курси вчення. Йому належать безліч публікацій на цю тему і відкриті звернення до чиновників від медичної науки. На сайті «Біометрика» зібрана обширна бібліотека книг і аналітичних публікацій, в розділі «Кунсткамера» проводиться

відкритий аналіз використання біостатистики в дисертаційних дослідженнях і ЖП. У тематичному випуску вищезгаданого журналу В.П. Леонов показує історичне коріння існуючої проблеми використання біостатистики в радянській клінічній науці і чому сучасна клінічна еліта блокує освоєння і впровадження сучасних міжнародних принципів представлення статистичного матеріалу в звітах КД, ЖП і дисертаціях. Виділяються також такі значимі чинники, як зниження виучуваності і студентів і фахівців, що склалися, інформаційний нігілізм медиків, роздвоєння інтересів вченого і медичного чиновника. Всі ці чинники перешкоджають освоєнню сучасної методології КД і якісному представленню статистичного матеріалу в ЖП. У плані обговорення альтернативних доріг подальшого розвитку В.П. Леонов стверджує: необхідно повсюдно в крупних університетах створювати лабораторії біостатистики; основна надія покладається не на підготовку лікарів-дослідників в області біостатистики, а на вчення математиків основам медичних знань і епідеміології; основна проблема полягає не в поліпшенні викладання статистики, а в забезпеченні якісного обов'язкового рецензування ЖП в редакціях медичних журналів і дисертацій в спецрадах. Звертається увага на багатолітні традиції лабораторій біостатистики в розвинених країнах і на гостру потребу в організації видання вітчизняного журналу з проблем біостатистики (Леонов В.П., 2006).

У іншому номері вказаного журналу обговорюється відкритий лист О.Ю. Ребровою, керівника лабораторії медичної інформатики ГУ Наукового центру неврології РАМН (раніше НДІ неврології). У цьому листі, окрім обговорення проблем підготовки фахівців, акцентується увага на головних причинах низької якості використання біостатистики у вітчизняних публікаціях. У листі заявляється, що основна проблема полягає в тому, що керівники наукових установ «не зацікавлені у високій якості науково-дослідницьких робіт, оскільки результати таких робіт частенько входять в протиріччя з атестаційною або адміністративною політикою директората, а інколи — і з фінансовою політикою». Тобто йдеться про системній проблемі в російській науці (Реброва О.Ю., 2007).

У відповідях на цей лист в основному обговорюється питання: «з якого середовища фахівців готувати біостатистиків, з лікарів або з математиків?». Звертає увагу замітка В.П. Леонова, де автор ще раз вказує дійсні причини, по яких правляча наукова еліта всіляко перешкоджає і перешкоджає впровадженню системи незалежного рецензування наукових публікацій. До тих пір, поки немає незалежної і прозорої системи оцінки якості ЖП і дисертацій, в науковому співтоваристві буде відсутній попит на фахівців з біостатистики. На думку експерта, «одним із

способів вирішення цієї проблеми, є систематичне інформування наукової громадськості і державних органів, в першу чергу ВАК РФ, РАМН, міністерств охорони здоров'я і освіти про дійсне положення справ в даній області» (Леонов В.П., 2007). Тобто пропонується вести систематичну організовану інформаційну роботу серед широкої наукової громадськості, зокрема на сторінках популярного «Міжнародного журналу медичної практики», силами знов створеної Суспільної експертної ради «Статистика в медицині і біології».

Принципово важлива ідея про те, що саме громадська організація професіоналів проводитиме незалежне статистичне рецензування. Це означає, що в Росії в умовах підвищення прозорості процесу академічної атестації, коли автореферати дисертацій, що захищаються, стають доступні будь-якому користувачеві мережі Інтернет (постанова уряду Російської Федерації № 277 від 20.04.2006 р. http://www.biometrika.tomsk.ru/vak_4) у дисертантів з'являється стійка мотивація забезпечити в своїй науковій праці якісне представлення статистичного матеріалу. Інакше незалежний відгук фахівця, направлений в спецраду, може змінити долю претендента.

Таким чином, у Росії втілюються передбачення відомого футуролога Е. Тоффлера, який в своїх моделях показав, як в постіндустріальному товаристві «третьої хвилі» вільні потоки інформації і знання вибивають ґрунт з-під ніг всякого роду бюрократії і змінюють структуру влади — влада переходить до людей знання (Тоффлер Е., 2004). На відміну від Росії, в українському науковому клінічному співтоваристві доки немає ні вільних потоків інформації, ні критичної маси організованих фахівців в області біостатистики, а звідси витікає існуюча структура влади, у тому числі і в науці.

ОБ'ЄКТИ АНАЛІЗУ

Як впливає з попереднього розділу, для реалізації змін на кращий необхідний прилюдний аналіз конкретних публікацій, оскільки просто говорити і писати про важке положення в клінічній медицині з методами біостатистики — це малопродуктивна діяльність, яку можна порівняти з організацією «бурі в стакані». Відповідно до концепції В.П. Леонова системні поліпшення якості статистичного аналізу можуть статися лише у тому випадку, коли автори наукових робіт знають, що їх науковий продукт може пройти незалежну прилюдну статистичну експертизу. Тепер в Росії, після того, як всі докторські дисертації, що захищаються, доступні на сайті ВАК Російській Федерації за 3 три місяці до захисту, неминує почнетися процес незалежного оцінювання якості цих робіт. Тим більше що абсолютно логічно для цього використовувати можливість такої громадської організації, як Суспільна експертна рада «Статистика в медицині і біоло-

гії».

Деякі читачі можуть заявити про те, що автореферати повинні лише відображати основні положення дисертації і не зобов'язані містити детальний цифровий матеріал КД. Але ми заперечимо на можливі думки тим, що навіть в сучасному структурованому резюме до ЖП відповідно до міжнародних стандартів потрібно представляти «основні результати (по можливості, конкретні дані та їх статистичну значущість)» (Весвітня асоціація медичних редакторів, 2005; www.icmje.org). Враховуючи, що автореферат дисертації — це набагато змістовніший звіт про КД, можна укласти, що в авторефератах обов'язково повинні вказуватися основні «конкретні дані та їх статистична значущість», у тому числі і детальний виклад методів статистичного аналізу. Це означає, що автореферат дисертації є зручним і репрезентативним об'єктом для оцінки якості аналізу біостатистики в КД.

До цих пір прилюдно незалежну експертизу ЖП і захищених дисертацій по біомедичних спеціальностях вів лише В.П. Леонов на сайті «Біометрики» в розділі «Кунсткамера». У цьому розділі представлена обширна колекція російських наукових експонатів, які повною мірою можна віднести до потворних явищ в науці. Слід зауважити, що В.П. Леонов в своїх аналітичних роботах фактично не виходить за рамки питань біостатистики, тоді як існує ще ціла низка проблемних запитань в наукових публікаціях, які можуть бути піддані формальній експертизі. До цих питань, окрім тих, які позначені в національних вимогах ВАК, також відносяться: якість дизайну КД, авторський вклад в наукове дослідження, конфлікт інтересів, порушення біомедичної етики, грубі методичні помилки при використанні конкретних діагностичних технологій, плагіат і інші елементи недобросовісної наукової практики.

Наш перший досвід прилюдної оцінки якості конкретних наукових робіт представлений в публікації 2005 р. на основі аналізу матеріалів крупної міжнародної наукової конференції з анестезіології (Воробйов К.П., 2005б). У 2006 р. в рамках наукової конкуренції і з метою аналізу якості певного методу клінічної діагностики нами створений сайт «Кунсткамера клінічної науки» (<http://www.kkn0.narod.ru>). На цьому сайті представлений аналіз матеріалів двох авторефератів докторських дисертацій за фахом 14.01.30 анестезіологія та інтенсивна терапія (Хріпаченко І.А., 2007; Постернак Г.І., 2008). Ці дисертації, на відміну від ЖП, пройшли багаторівневу експертизу в установі-виконавцеві, спеціалізованій вченій раді і у офіційних опонентів. Тому є підстави вважати дані автореферати типовими екземплярами дисертацій вищого академічного рівня, що захищаються в Україні. Виходячи з цього

припущення, окрім загальних критичних зауважень до типових статистичних помилок, ми використовуватимемо матеріал цих двох робіт, які позначені як експонати № 1 (http://www.kkn0.narod.ru/P2/H1_Main.html) і № 3 (http://www.kkn0.narod.ru/P4/H3_Main.html)

ТИПОВІ ПОМИЛКИ АНАЛІЗУ БІОСТАТИСТИКИ

Спроба опису відомих помилок статистичного аналізу в КД займе багато томів, оскільки, на відміну від єдиного вірного принципу використання статистичного критерію, існує безліч найрізноманітніших помилок його вживання. Тому в рамках подібної аналітичної публікації є лише одна можливість — класифікувати типів найбільш очевидних і поширених помилок і коротко описати типові помилки вживання методів біостатистики з використанням прикладів.

Під поняттям «Типові помилки» маються на увазі певні класифікаційні типи помилок представлення результатів статистичного аналізу. Поняття «Типові помилки» розглядається як помилки, що належать до певного класифікаційного типу. Тобто для аналізу типових помилок спочатку слід визначити класифікацію типів. У даній публікації ми пропонуємо розглянути помилки, які зустрічаються при:

- описі методів статистичного аналізу;
- представленні вибіркового характеру;
- зіставленні двох незалежних груп;
- проведенні множинних порівнянь незалежних вибірок;
- аналізі зв'язаних вибірок;
- описі результатів діагностичних досліджень;
- помилках, пов'язаних з невірною візуалізацією результатів статистичного аналізу.

ЩЕ РАЗ ПРО РОЛЬ ДИЗАЙНУ КД

Перш ніж переходити до статистичного аналізу клінічного матеріалу, необхідно переконатися в тому, що цей матеріал адекватно відображає популяцію, що вивчається. У попередній публікації вказано, що основне завдання якісного дизайну КД — це мінімізація систематичних помилок (Воробйов К.П., 2008б). Для того, щоб краще представити поняття «Систематична помилка», а також цифрові наслідки невірної дизайну КД, ми на основі відомих образів створили чотири моделі збору вибіркового характеру для популяції, в якій ознака, що вивчається, зустрічається відповідно до нормального закону розподілу (рис. 1).

Перша модель (А) характерна для ідеально спланованого дизайну, в якому відсутні систематичні помилки і якісно розроблені критерії включення/виключення в/з дослідження. Для такої моделі також характерний низький рівень вірогідності випадкових помилок, що виявлятиметься відносно низькою вибірковою дисперсією.

єю, а в результаті розрахунку статистичних характеристик в цій моделі середні значення відповідатимуть середнім в популяції.



Рис. 1. Чотири моделі даних, відібраних в одній популяції при різному дизайні КД, яким відповідають різні репрезентативність і різний статистичний результат. Синім кольором позначена гістограма частоти тієї, що зустрічається ознаки (вертикальна вісь), що вивчається, залежно від вираженості прояву цієї ознаки (горизонтальна вісь) в популяції, що вивчається (генеральній сукупності). А — високоселективна вибірка і низькі систематичні помилки; В — низькоселективна вибірка і низькі систематичні помилки; С — високоселективна вибірка, але наявність систематичної помилки; D — низькоселективна вибірка і наявність систематичної помилки.

Друга модель (В) за всіх інших рівних умов може відрізнитися від попередньої нижчою якістю відбору хворих в дослідженні. Це може бути пов'язано з тим, що не вироблялися обмеження відбору хворих в дослідження по декількох чинниках, що втручаються, які в моделі (А) були враховані, або не враховано декілька джерел інших систематичних помилок, причому ці додаткові чинники надають різноспрямовану дію на вимірюваний результат. У цій моделі вибірккові середні також відповідають середньою популяцією, але із-за вищої вибіркової дисперсії довіра до результатів дослідження в цій моделі збору даних буде нижче.

Моделям С і D за інших рівних умов відповідають високі, але різноспрямовані систематичні помилки. Наприклад, в цих моделях низька вираженість ознаки (С), що вивчається, могла бути обумовлена включенням в дослідження людей молодшого віку, а більш виражені прояви ознаки в іншій моделі (D) — включенням в дослідження людей з наявністю певної шкідливої звички (куріння). Ні вік, ні спосіб життя не вивчалися в такому гіпотетичному дослідженні, але вони могли істотно вплинути на результат і дати зміщену оцінку по відношенню до популяції. Тобто ці дві моделі даних ні за яких умов не дозволяють зробити вірні висновки про характеристики популяцій, що вивчаються. Значить, який-небудь статистичний аналіз даних, отриманих в аналогічних випадках, не має жодного сенсу.

На основі цієї моделі представимо ситуацію, в якій зіставляються три по-різному організованих КД, що оцінюють певний чинник ризику в одній і тій же популяції. Відповідно до дизайну кожного з цих КД отримані дані, аналогічні моделям С, А і D. Виходячи з даних (див. рис. 1)

вочевидь, що при формальному статистичному зіставленні цих вибірок буде зроблений висновок, що ці три групи даних відносяться до різних генеральних сукупностей. Цей помилковий висновок буде статистично бездоганний. Даний приклад підтверджує відомий вислів А.Ф. Johnson (1985) про те, що «при істотних недоліках в плануванні або виконанні дослідження не допоможе жоден формальний статистичний аналіз» (Флетчер Р. і співавт., 1998). Таким чином, перш ніж оцінювати статистичний матеріал в яких-небудь публікаціях, слід уважно вивчити опис дизайну КД. По рекомендаціях експертів-епідеміологів, якщо в публікації неоднозначно описані умови формування груп, що вивчаються, то такі роботи не слід сприймати серйозно — вони годяться лише для смітцевої корзини (Власов В.В., 2001; Грінхальх Т., 2004).

В авторефераті експонату № 1 представлено 118 числових вибірккових характеристик (середнє і стандартне відхилення), в яких коефіцієнт варіації >30%, що свідчить про надмірно великий розкид клінічних даних (див. нижче). При такому описі клінічного матеріалу існує висока вірогідність систематичної помилки, пов'язаної з дизайном даного КД.

Нещодавно (2007) вийшла ґрунтовна публікація відомого епідеміолога В.В. Власова, в якій представлений аналіз причин виникнення систематичних помилок і способи їх контролю (Власов В.В., 2007).

ОПИС МЕТОДІВ СТАТИСТИЧНОГО АНАЛІЗУ

У цьому питанні зазвичай посилаються на міжнародні рекомендації редакторів біомедичних журналів (Всесвітня асоціація медичних редакторів, 2005; www.icmje.org). У десятці рядків розділу «Статистика» цих рекомендацій виділені головні вимоги для авторів, але їх не можна визнати досить повними. З інших відомих аналогів найбільш повніші і якісніші рекомендації за представленням статистичного матеріалу пропонує загальновідоме рейтингове періодичне видання «Nature medicine» (Guide to authors, 2008). Ці статистичні рекомендації (СР) маловідомі вітчизняному дослідникові, тому ми пропонуємо познайомитися з останнім оновленням (10.04.2008) СР в нашому перекладі.

«Статистичні рекомендації»

Кожна стаття, в якій використовуються статистичні методи, повинна вказувати найменування статистичного тесту, кількість (n) в кожному статистичному аналізі, обґрунтування вибору певного тесту (включаючи обговорення нормальності розподілу, якщо тест призначений лише для нормально розподілених даних), а-рівень для всіх тестів, були тести односторонніми або двосторонніми і фактичні значення «Р» для кожного тесту (не просто «істотно» або «р<0,05»). Повинно бути ясно вказано, який ста-

тистичний тест використаний для здобуття даного р-значення. В разі коротких повідомлень ці відомості повинні повідомлятися в тексті або відбиватися на малюнках і в підписах до них.

Вибіркові дані мають бути представлені у вигляді описових статистик, які повинні включати: розміри вибірки (n) для кожного набору даних; заходи розсіяння, такі як стандартне відхилення або розмах. Для малих вибірок розмах є прийнятнішим, ніж стандартне відхилення. На графіках мають ясно відображувати кордони помилок (error bar — «вуса») або довірчих інтервалів. Автори зобов'язані вказувати, чи є числа, що йдуть після знаку \pm , стандартними помилками середнього (s.e.m.) або стандартним відхиленням (s.d.).

Автори зобов'язані обґрунтувати вживання конкретного критерію і пояснити, чи узгоджуються аналізовані дані з допущеннями для тесту. Найбільш поширеними є наступні три помилки:

Численні порівняння. Коли на одному наборі даних проводяться множинні порівняння, автори зобов'язані вказати, як вони адаптували α -рівень для обліку помилки першого типу або необхідно використовувати статистичні критерії, призначені для численних порівнянь (такі як ANOVA, а не серія t-тестів).

Нормальність розподілу. Багато статистичних критеріїв вимагають, аби дані були розподілені приблизно нормально; використовуючи ці тести, автори повинні вказати, як вони перевірили свої дані на нормальність. Якщо дані не відповідають умовам вживання тесту, то повинна використовуватися непараметрична альтернатива.

Малий об'єм вибірки. Коли об'єм вибірки малий (<10), то автори зобов'язані використовувати критерії, придатні для аналізу малих вибірок, або ж обґрунтувати правомірність використання тестів для великих вибірок

Контрольний список для мінімізації ризику статистичних помилок доступний за адресою: http://www.nature.com/nm/authors/submit/Checklist_of_statistical_adequacy.doc;

Для нашого дослідження також актуальні наступні рекомендації цього документа з розділу «Підготовка графіків»: графіки мають бути на білому фоні, слід уникати рамок, непотрібних кольорів, декоративних ефектів (таких як тривимірні графіки), малюнків з високим дозволом. Вертикальна вісь гістограми не має бути усіченою, щоб перебільшувати незначні відмінності».

У поточних клінічних публікаціях найчастіше взагалі не описують ні сенс виразів типу $A \pm b$, ні яким статистичним тестам відповідає рівень значущості «р» (якщо він вказується). Часто замість вказівки граничного рівня α -помилки і точних значень «Р» в публікаціях йдеться про «достовірність відмінностей» з вказівкою однієї-трьох зірочок поряд з вибірковими характеристиками,

що зіставляються. У чотирьох експозиціях «Кунсткамери» В.П. Леонова можна знайти найдивніші описи методів статистики в журнальних публікаціях і дисертаціях, але, вочевидь, все-таки приз «Кунсткамери» повинен належати нашому вітчизняному експонату № 3: у авторефераті докторської дисертації в пункті «Методи дослідження» заявлено коротко і ємно «статистичні методи». Більше слово статистика і пов'язані з ними статистичні критерії або які-небудь розшифровки цифрових позначень в авторефераті не згадуються.

ПОМИЛКИ ПРЕДСТАВЛЕННЯ ВИБІРКОВИХ ХАРАКТЕРИСТИК

Цей тип помилок пов'язаний з незнанням дослідниками найбільш простих правил статистики, а тому їх можна позначити як грубі помилки. Найбільш важливим і самодостатнім матеріалом будь-якого КД є вибіркові числові характеристики об'єктів дослідження, до яких найчастіше відносяться « межі положення » і « межі розсіяння » (Петрі А., Себін Д., 2003). Ці показники дозволяють оцінити центральні тенденції і розкид показника у вибірці, що вивчається. Для грамотного фахівця ця інформація дозволяє висловити власні думки про якість клінічного матеріалу. Основна і найбільш поширена помилка в поточних публікаціях — це представлення в публікаціях виразів типу $A \pm b$, сенс яких часто не розкривається. Якщо значення знаку «А», як правило, не викликає особливих питань, то $\pm b$ може позначити принаймні три показники міри розсіяння: стандартне відхилення, стандартну помилку середнього і коливання довірчого інтервалу. Окрім цього, дуже рідко автори поряд з виразами $A \pm b$ дають інформацію про розмір вибірки (n), знання якої необхідні для зворотного обчислення характеристики міри розсіяння (якщо $\pm b$ є оцінкою точності середнього) або для обчислення точності оцінки вибіркового середнього (якщо $\pm b$ є характеристикою дисперсії). І лише одиничні публікації представляють вибіркові характеристики з вказівкою 95% довірчого інтервалу, тоді як епідеміологи заявляють, що «Визначення довірчих інтервалів стало звичайним способом представлення головних результатів клінічних досліджень, оскільки такий підхід володіє багатьма перевагами перед методом перевірки гіпотез (оцінкою р)» (Флетчер Р. і співавт., 1998, с. 246).

Друга група проблем пов'язана з описом в публікаціях вибірок, не відповідних нормальному закону розподілу. Опис таких вибірок за допомогою середніх і показників дисперсії не відображає реальних уявлень про характер розподілу ознаки, що вивчається, у вибірці. Для цих вибірок використовують медіано-квартильні характеристики. Принаймні дослідник при виборі альтернативи повинен приблизно оцінити розподіл даних в експериментальній вибірці. Покажемо на

спеціальному прикладі, до чого може привести ігнорування вищезгаданих правил. Для демонстрації матеріалу публікації ми створили два наступні цілочисельні ряди з однаковими розмірами вибірок ($n=21$).

Ряд 1: 21, 22, 22, 23, 23, 24, 24, 24, 24, 25, 25, 25, 25, 26, 26, 26, 26, 27, 27, 28, 29, 30.

Ряд 2: 20, 20, 20, 20, 20, 20, 21, 21, 21, 21, 21, 21, 21, 21, 21, 33, 34, 34, 36, 37, 42.

При подальшому викладенні матеріалу ми використовуватимемо наступні позначення статистичних характеристик:

M — середнє арифметичне; M_o — мода; M_e

— медіана; SS — дисперсія; S — стандартне відхилення; m — стандартна помилка середнього; $95\%M$ — 95% довірчий інтервал розсіювання середнього; $0,5L$ — нижній квантиль; $0,5U$ — верхній квантиль; V — коефіцієнт варіації; n — розмір вибірки.

Чудовим фактом є приблизна рівність середніх цих двох вибірок ($M_1=25,14$; $M_2=25,00$), але висока відмінність меж розсіювання ($S_1=2,31$; $S_2=7,32$) пов'язана з різним розподілом числових рядів: перший ряд підкоряється нормальному закону розподілу, другий — істотно відрізняється від нормально розподілених даних (рис. 2).

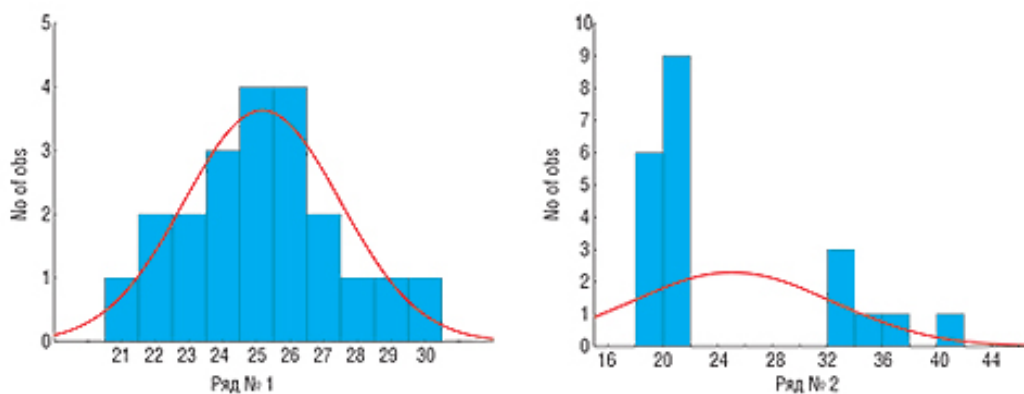


Рис. 2. Гістограми розподілу числових рядів, що вивчаються.

Вважатимемо, що ці два ряди є випадковими вибірками з двох генеральних сукупностей. Можливі варіанти числового опису вибірових характеристик цих двох рядів даних представлені в наступній таблиці (таблиця).

При описі першої вибірки, яка підкоряється нормальному закону розподілу, достатньою інформацією буде повідомлення про $M \pm S(n)$. Ще правильнішою буде вказівка середнього значення, верхнього і нижнього 95% довірчих інтервалів для генеральної середньої, які розраховуються як $M - 1,96 \cdot m$ і $M + 1,96 \cdot m$. Позначення розміру вибірки (n) в кожному випадку опису вибірових характеристик суворо рекомендовано відомими міжнародними стандартами.

Клінічна інформація не завжди розподіляється по нормальному закону, більш того, як помічають епідеміологи: «Схожість реальних розподілів з кривою нормального розподілу в клінічній практиці, як правило, випадково» (Флетчер Р. і співавт., 1998). При розподілі даних згідно із законом, що відрізняється від нормального, середнє і довірчі інтервали не дають повного уявлення про вибірку. Тому для характеристики другої вибірки слід використовувати M_e і квантильні оцінки. Звертаємо увагу на той факт, що медіана, нижній і верхній квантиль для першої вибірки фактично відповідають її середньому і кордонам відповідного 95% довірчого інтервалу, тоді як медианно-квантильні оцінки другої ви-

бірки істотно відрізняються від її опису через M ($M - 1,96 \cdot m$; $M + 1,96 \cdot m$) (див. таблицю.). Цей факт є наочною демонстрацією необхідності опису нерівномірно розподілених даних через медіано-квантильні оцінки.

Для вибору способу характеристик числових рядів (M або M_e) може бути корисне відображення гістограми розподілу вибірки або результату формального тесту аналізу розподілу. Наприклад, за даними тесту Шапіро — Уїлки прийняти нульову гіпотезу про приналежність вибірок до нормально розподілених даних можна для першої вибірки з вірогідністю 0,89, для другої — $< 0,00001$. Тобто другу вибірку слід описувати і застосовувати до неї тести як до вибірки, що відрізняється від нормального закону розподілу. Ще однією, але менш прийнятною характеристикою нормальності розподілу, є коефіцієнт варіації $V = S / M \cdot 100\%$ ($V_1 = 5\%$; $V_2 = 54\%$). Вважається, що «якщо $V < 30\%$, то ряд даних не має значної скісності», а якщо $V > 100\%$, то це означає, що «дані неоднорідні» (Лапач С.Н. і співавт., 2002). В іншому випадку епідеміолог визначає V як «...безрозмірну величину, менше одиниці» (Власов В.В., 2005). Тобто в нашому випадку при настільки високому V_2 невірно описувати вибірку через середні величини і характеристики дисперсії.

Поглянемо на опис вибірових характеристик в аналізованих авторефератах. У експонаті №

1 автор описує вибірки через характеристики $M \pm S$, при цьому в 118 випадках опису вибірко-вих середніх $V > 33\%$. На гістограмі (рис. 3) представлена частота тієї, що зустрічається V залежно від його величини.

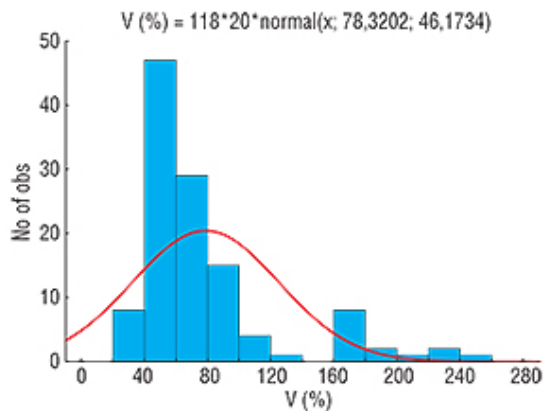


Рис. 3. Розподіл 118 коефіцієнтів варіації вибірко-вих характеристик в матеріалі автореферату (експонат № 1, включені $V > 33\%$).

У 19 випадках $S > M$, тобто коефіцієнт варіації в цих вибірках $> 100\%$, а в 4 випадках $V > 200\%$. Зіставивши наш приклад (Ряд № 2, див. рис. 2, $V_2 = 53\%$), читач може представити, наскільки виражена неоднорідність вибірок у вказаному авторефераті. При цьому слід зауважити, що йдеться про основний масив клінічного матеріалу, який відбитий в авторефераті. В авторефераті автор вказав на необхідність оцінки нормальності розподілу вибірок, але в тексті автореферату при уявленні фактичного матеріалу про нормальність розподілу не згадується. Очевидні відхилення розподілу від нормального в більшості випадків вимагали роз'яснення та іншого формату уявлення заходів положення і розсіяння через медіанні оцінки. Також при описі вибірко-вих характеристик не наводяться поряд із значеннями $M \pm S$ розміри вибірок (окрім таблиць), що утрудняє сприйняття матеріалу.

Другий автореферат (експонат №3) є унікальним за формою представлення первинного матеріалу: автор взагалі не показує яких-небудь вибірових характеристик результатів дослідження в абсолютних значеннях, а використовує часткове представлення результатів, наприклад: «Концентрація ... зростала впродовж першої доби спостереження на $50,8 \pm 4,5\%$, а в бронхіальному секреті в $19,7 \pm 4,1$ разів». Існує поняття довірчий інтервал для частки, але такий інтервал вказується у вигляді діапазону з двох цифр з поясненням способу його обчислення та сенсу цього діапазону. Вищезгаданий приклад представлення характеристики частки без вказівки сенсу числових значень після знаку \pm по суті викладу є абсурдом, але цей абсурд пройшов багаторічну експертизу в системі академічної атестації, що

діє.

ЗІСТАВЛЕННЯ ДВОХ НЕЗАЛЕЖНИХ ГРУП

Цей тип помилки вважається найбільш поширеним. Основна помилка зв'язана з використанням критерію Стьюдента без перевірки умови можливості його вживання. Велика частина екземплярів в «Кунсткамері» В.П. Леонова присвячена аналізу цієї типової помилки. Продемонструємо на наших числових рядах проблему застосовності критерію Стьюдента для двох незалежних вибірок. При перевірці нульової гіпотези для першого і другого рядів вищепредставлених даних ми отримуємо $p = 0,93$. Тобто згідно з цим результатом з вірогідністю 93% можна стверджувати, що середні значення двох досліджуваних груп однакові і взяті з однієї генеральної сукупності. Проте у вказаному випадку ми порушуємо допустимість використання критерію Стьюдента, оскільки ряд № 2 розподілений згідно із законом, який відрізняється від нормального і між дисперсіями рядів існують істотні відмінності ($SS_2 > SS_1$, F-критерій Фішера $= 10,1$; $p < 0,000003$). Тепер виконаємо коректне зіставлення цих вибірок за допомогою непараметричного тесту. За даними тесту Манна — Уїтні можна лише з вірогідністю 2,3% стверджувати, що середні величини та медіани вибірок, що зіставляються, належать одній генеральній сукупності. Цей приклад наочно показує, які невірні висновки можуть бути зроблені при неправильному використанні критерію Стьюдента.

Зіставляючи спосіб перевірки гіпотез за допомогою значення «р» і методів довірчих інтервалів, епідеміологи звертають увагу на безперечні переваги другого способу: «Цей підхід став широко застосовуватися нещодавно і зараз віддається перевага більшістю журналів по причинах, розглянутих нижче» (Флетчер Р. і співавт., 1998). Суть цього методу проста — якщо довірчі інтервали двох порівнюваних вибірок не перекриваються, тоді відмінності вважаються встановленими відповідно до рівня довірчого інтервалу. Якщо поглянути на дані таблиці, то стає очевидним, що довірчі інтервали рядів № 1 і № 2 перекриваються, але це ще не означає (як показано вище), що дані відносяться до однієї генеральної сукупності. Довірчі інтервали годяться лише для обґрунтування припущення про відмінність вибірок, а не про їх рівнозначність.

Аналогічні положення поширюються на інтерпретацію величини «р». При зіставленні двох незалежних груп перевіряється так звана нульова гіпотеза, тобто вірогідність приналежності вибірок до однієї генеральної сукупності — це і є величина «р». Інколи, якщо величина «р» більше обумовленого критичного значення, деякі автори роблять висновок про те, що між параметрами, що вивчаються, немає відмінностей. Це невірний висновок, і ось що з цього приводу пишуть фахі-

вці: «Висновок про те, що між способами втручання немає відмінностей, оскільки p перевищує деякий рівень — некоректний; у цій ситуації слід застосовувати показник p (вірогідність b -помилки)» (Флетчер Р. і співавт., 1998). «На жаль, обґрунтування висновків прийняттям нульової гіпотези — дуже поширена помилка в медичних дослідженнях» (Власов В.В., 2001). «Статистична потужність є вірогідністю виявлення статистично значимої відмінності за умови, що воно дійсно існує. Статистично незначимі результати, отримані в дослідженні з малою статистичною потужністю, невірно вважати негативними; вони недостатні: «відсутність гарантії не є гарантія відсутності» (Ланг Т., 2005).

Покажемо помилки зіставлення двох незалежних груп на прикладі. У експонаті № 1 на с. 7 автореферату в описі статистичних методів говориться, що «непараметричні критерії порівняння використовувалися для змінних, при оцінці яких гіпотеза про нормальність розподілу відкидалася». Але при цьому нічого не вказується про другу, важливішу умову вживання параметричних критеріїв — перевірку рівності дисперсій. Статистичні підручники стверджують, що критерій «менш стійкий при нерівних дисперсіях», ніж при відхиленнях від нормальності розподілу (Петрі А., Себін Д., 2003). Тобто перевірка рівності дисперсій при використанні критерію Стюдента є найбільш важливою умовою його вживання.

При аналізі цього автореферату в деяких випадках нам удалося відновити необхідні дані (n) для зіставлення дисперсій вибірок, що вивчалися: на с. 26 використовується критерій Стюдента, при порівнянні двох вибірок по формату $M(S; n)$ вибіркові характеристики відповідають 3,28 (2,3; $n=45$) і 7,7 (3,44; $n=75$). Відношення дисперсій в цих вибірках дорівнює 2,4, що відповідає вірогідності рівності дисперсій 0,003. Тобто дисперсії не рівні і критерій Стюдента в даному випадку використовувати недоречно. На с. 27, третій абзац, для пари вибірок з характеристиками 7,21 (0,09; $n=70$) і 7,22 (0,07; $n=57$) вірогідність рівності дисперсій складає 0,026, що також не дозволяє використовувати критерій Стюдента. Якби автор скрізь при описі вибіркової характеристики і їх порівнянні вказував використовуваний статистичний критерій і розміри вибірок, то можна передбачити, що були б виявлені більш багаточисельні помилки. Тобто невірний формат представлення статистичних результатів не дозволяє зробити однозначної оцінки якості більшості статистичних результатів в авторефераті. Відповідно до загальноприйнятого принципу «презумпції винності» ученого в представленні наукових результатів кожен дослідник зобов'язаний давати однозначну і точну інтерпретацію свого фактичного матеріалу, а інакше його матеріал не повинен прийматися науковим

співтовариством. Інший стан речей слід віднести до системних проблем в науковому співтоваристві.

Інколи автори невірно інтерпретують обчислений рівень статистичної значущості « p ». Це частіше стосується використання слів «достовірне», замість «значимо», що свідчить про нерозуміння дослідником ества статистичного аналізу, що проводиться. У експонаті № 1 на с. 26 автореферату автор так інтерпретує досягнуте р-значення: «... пацієнти першої вибірки характеризувалися статистично значимо вищими значеннями... Вірогідність безпомилкової думки склала 0,049». Тобто автор стверджує, що довіряє своїм думкам лише на 4,9%. У іншому випадку на с. 19, визначивши статистичну значущість для коефіцієнта кореляції $p < 0,05$, автор стверджує, що «Біля померлих коефіцієнт Пірсона був достовірний в 0,95% довірчому інтервалі, і дорівнював 0,52», при цьому сам довірчий інтервал не наводиться. Вочевидь, автор хотів сказати про 95% упевненості в тому, що якийсь довірчий інтервал (мають бути приведені дві цифри, які позначають діапазон) правильно відображає обчислений коефіцієнт кореляції. У такій ситуації виникають сумніви в розумінні претендентом сенсу сказаного, оскільки неможливо говорити про яку-небудь «достовірність» при 0,95% довірі до результату обчислення і неможливо оцінити інтервал по величині однієї цифри. У іншій фразі на цій же сторінці автор пише: «відмінності між групами достовірні в 88% довірчому інтервалі», але при цьому ніяк не визначає цей інтервал. Можна передбачити, що 88% довірчих інтервалів двох порівнюваних вибірок не перетинаються. Але для такого затвердження необхідно привести значення цих інтервалів, а цього немає в авторефераті (дані представлені у форматі $M \pm S$). Вочевидь, для порівнюваних вибірок обчислений рівень статистичної значущості відмінностей $p = 0,12$ (за результатами неоголошеного критерію при коефіцієнті варіації вибірок 38 і 37%). Якщо використаний адекватний тест, то результат означає, що в 12% випадків порівнювані генеральні середні величини будуть рівні, але це не дозволяє стверджувати, що в 88% випадків вони різні. Тобто, судячи по тексту автореферату, автор змішує різні поняття, оскільки довірчий інтервал задається дослідником довільно, а потім обчислюється і позначається діапазоном з двох цифр.

МНОЖИННІ ПОРІВНЯННЯ НЕЗАЛЕЖНИХ ВИБІРОК

Проблема множинного порівняння незалежних груп найчастіше з'являється в пошукових дослідженнях, коли автор, не висуваючи первинної гіпотези, після закінчення збору матеріалу починає зіставляти можливі пари спостережень за допомогою критерію Стюдента. Цій проблемі приділяється досить багато уваги у всьому ста-

тистичному керівництві, але, проте, автори до цих пір рідко використовують в цих випадках дисперсійний аналіз (ANOVA). Звернемося до аналізу конкретного автореферату (експонат № 1). На с. 26 автор зіставляє попарно 3–4 і 3–5 груп і використовує для цього критерій Стюдента. Таке попарне зіставлення вимагає використання спеціальних поправок для обчислених р-значень, тобто висновки претендента по цих обчисленнях невірні.

АНАЛІЗ ЗВ'ЯЗАНИХ ВИБІРОК

У КД часто використовується так званий аналіз даних до–після, наприклад при зіставленні певного параметра у одних і тих же хворих в різні проміжки часу до та після дії чинника, що вивчається. Такі групи називають зв'язаними, а дані парними. Використання стандартного критерію Стюдента для зіставлення зв'язаних груп є грубою помилкою. Для цих цілей використовуються спеціальні статистичні методи, наприклад критерій Стюдента для парних порівнянь, критерій знакових рангів Вілкоксона. У випадку якщо порівнюються три і більш зв'язані вибірки, використовується дисперсійний аналіз повторних вимірів. Ми вже показували цю помилку на матеріалах міжнародного конгресу анестезіологів і на моделі автора однієї з робіт продемонстрували, що невірний підхід в аналізі до–після може істотно спотворити дійсні результати дослідження (Воробйов К.П., 2005б). В авторефераті експоната № 1 в таблицях на с. 12 і 13 говориться про зіставлення відмінностей до–після. Не відомо, який статистичний метод для цього використовується, оскільки в переліку вживаних статистичних тестів не вказаний тест для парних вибірок.

ОПИС РЕЗУЛЬТАТІВ ДІАГНОСТИЧНИХ ДОСЛІДЖЕНЬ

Один з найбільш методично складних статистичних підходів використовується в сучасних КД, які присвячені вивченню методів діагностики. Для таких досліджень існує певний стандарт, який, судячи по поточних публікаціях і дисертаціях, не відомий вітчизняним дослідникам. Класичний дизайн діагностичних досліджень передбачає зіставлення референтного методу діагностики (золотого стандарту діагностики) з діагностичним параметром, що вивчається, у одних і тих же пацієнтів. Потім спеціальним чином будується так звана характеристична крива, яка є операційною характеристикою діагностичного тесту та визначається точка розділення, яка розділяє хворих і здорових. Після цього складається так звана «чотирипільна таблиця» і на основі її даних обчислюються чутливість, специфічність і інші числові характеристики нового діагностичного тесту. Такий підхід дозволяє отримати певні статистичні характеристики і має безумовні переваги — результати дослідження стають абсолютно зрозумілими і однозначними. Інші методичні

підходи в діагностичних дослідженнях без відповідного обґрунтування слід розглядати як грубу методологічну помилку. Детальний виклад методології діагностичних досліджень викладений в серії публікацій В.В. Власова, зокрема в публікації 2006 р. (Власов В.В., 2006б).

ВІЗУАЛІЗАЦІЯ РЕЗУЛЬТАТІВ СТАТИСТИЧНОГО АНАЛІЗУ

Неякісна візуалізація часто має завдання маніпулювання думкою читачів. Особливо це справедливо по відношенню до різних популярних рекламних медичних видань. Тому часто помилки візуалізації статистичних даних слід розглядати з точки зору умисної помилки. Приклад такого маніпулювання думкою читача наведений в нашій публікації (Воробйов К.П., 2005а).

Для різних статистичних методів аналізу використовуються відповідні прийоми візуалізації. При представленні вибірових характеристик сучасним стандартом є так званий графік «ящик з вусами», який при нормальному розподілі досліджуваної вибірки може містити інформацію про середню величину, стандартну помилку та 95% довірчі інтервали середньої. При розподілі, що відрізняється від нормального, такий графік повинен відображати інформацію про значення медіани, квантилях (процентиліях) і характеристиках викидів або про 95% довірчих інтервалів. На прикладі наведених вище числових рядів продемонструємо переваги кожного з цих підходів (рис. 4).

Візуалізація вибірових характеристик за допомогою графіка «ящик з вусами», в якому вуса представляють 95% довірчий інтервал, є сучасним методичним стандартом при зіставленні двох нормально розподілених вибірок. Якщо верхній кордон вусів одного графіка не перекриває нижнього кордону іншого графіка, то такі вибірки можна з упевненістю віднести до різних генеральних сукупностей. Також поважно відзначити, що на відміну від величини «р» кордону вусів дозволяють візуалізувати величину розкиду ознаки (наприклад клінічного ефекту). Якщо вибірки не відповідають нормальному закону розподілу, тоді медіанно-квантильні графіки дозволяють отримати точніші уявлення про ряди даних, що вивчаються. Якщо при використанні на графіці середніх і 95% довірчого інтервалу зіставляється вибірка з асиметричними даними, то ховається структура вибірок, що вивчаються. Відображення вибірових характеристик цих же числових рядів за допомогою медіанно-квантильних оцінок дозволяє виявити особливості розподілу даних у вибірках (див. рис. 4).

У авторефераті експоната № 1 на с. 26 дисертант доводить одну з ключових закономірностей свого дослідження, зіставляючи рівень лактату крові хворих з різним прогнозом. У групі хворих, що вижили, рівень лактату складав $4,54 \pm 3,52$ ммоль/л ($M \pm S$), а в групі померлих —

7,88±3,37 ммоль/л. Коефіцієнти варіації в цих випадках склали відповідно 77 і 43%, що вказує на неоднорідність і високу вірогідність порушення нормального розподілу цих вибірок. Автор не лише невірною використував критерій Стьюдента для зіставлення середніх цих вибірок (без перевірки нормальності розподілу), але також застосував для цих явно асиметричних даних графіки «ящик з вусами», в яких параметрами служили середні величини і 95% довірчий інтервал. Цілком можливо, що дані дійсно можна

віднести до різних генеральних сукупностей, але медіанні оцінки і непараметрична статистика в цьому випадку дозволили б авторам коректніше представити структуру розподілу даних у вибірках і довести відмінності середніх величин і медіан. Читач може зіставити вищезгадані коефіцієнти варіації і графік медіанної оцінки ряду № 2 на рис. 4, який має виражену асиметрію при нижчій неоднорідності вибірки ($V^2=53\%$), чим в аналізованому авторефераті.

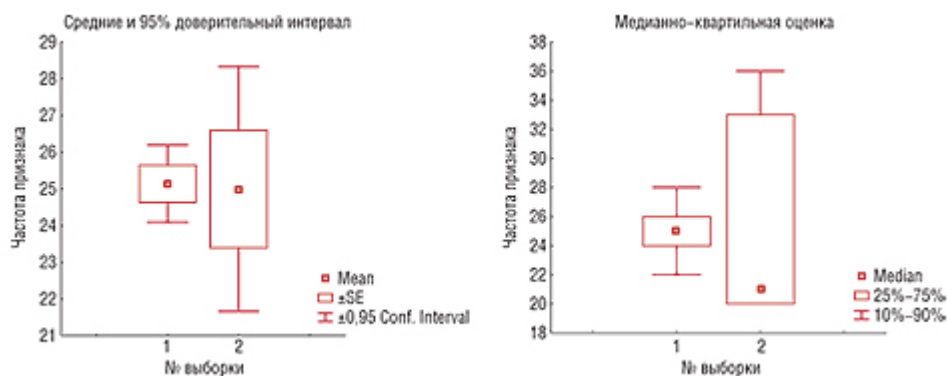


Рис. 4. Два альтернативні підходи візуалізації порівняльної точкової оцінки рядів даних № 1 і № 2.

При графічному зіставленні двох зв'язаних вибірок рекомендується використовувати графіки, на яких значення показника до–після, що вивчається, з'єднуються прямою лінією уздовж осі абсцис. В разі використання стовпчастої діаграми середніх для відображення середніх значень до–після можуть бути загублені різноспрямовані тенденції динаміки показників у зв'язаних групах, як це було показано нами на прикладі в спеціальній публікації (Воробйов К.П., 2005б).

Правильна візуалізація результатів діагностичного дослідження дозволяє по величині площі під характеристичною кривою отримати пряму оцінку ефективності діагностичного тесту, що вивчається.

Слід також зробити декілька зауважень про засоби побудови наукових графіків і програми, які для цього можуть використовуватися. У виступах на конференціях і в публікаціях дуже часто використовуються барвисті, тривимірні, конусоподібні та інші графіки, але всі ці прикраси не несуть якої-небудь інформації про характеристики розкиду та об'єми вибірок. Такі графіки приховують дійсні результати досліджень і можуть розглядатися як засіб маніпуляції думкою аудиторії і читачів. Кожен елемент прикраси графіка зобов'язаний нести, перш за все, змістовну інформацію. Також слід зауважити, що популярна програма Excel має обмежені можливості візуалізації наукової інформації і в більшості випадків не дозволяє якісно відображувати результати сучасних статистичних методів досліджень.

ВИСНОВОК

Існує ще більша група помилок вживання статистичного аналізу в ЖП і дисертаціях, які пов'язані з рідко використовуваними методами статистичного аналізу або із складними статистичними підходами (багатофакторний аналіз, кластерний аналіз, аналіз дискримінанта). Також ми не торкнулися використання статистичного аналізу для якісних даних. Всі ці питання детально викладаються у ряді спеціального керівництва і аналізуються в тематичних публікаціях фахівців. В рамках даної публікації ми лише хотіли затвердити існування проблеми, яка наполегливо ігнорується нашим науковим співтовариством, не дивлячись на достатній об'єм доступної інформації. У Росії приходить розуміння проблем і бачення доріг, їх рішення. Ось що із цього приводу пише головний редактор «Міжнародного журналу медичної практики» В.В. Власов у вступному слові до тематичного випуску, який присвячений вченню статистиці: «На жаль, з роками ситуація не покращується, не дивлячись на те що проблема в цілому зрозуміла, і ясно, в якому напрямі треба рухатися ... привласнення вчених ступеней відбувається, не дивлячись на неспроможність статистичного аналізу, а в наукових журналах рецензування статей відбувається силами осіб, що знають про статистику зазвичай з чуток». Далі говориться про вирішення Суспільства фахівців доказової медицини: «активізувати роботу Суспільства у напрямі критики методологічних помилок медичних досліджень, перш за все в медичних дисертаціях ... для того, щоб створити необхідний тиск на відс-

талу систему» (Власов В.В., 2006а).

У даній публікації представлений досвід аналізу методологічної якості матеріалів авторефератів двох докторських дисертацій. Результати цієї роботи підтверджують існування прихованої проблеми. Причому якщо в експонаті № 1 дисертант спробував застосовувати всілякі статистичні підходи і зробив ряд помилок, то в експонаті № 3 дисертант фактично не наводить статистичних даних, допускає грубі помилки при викладі цифрового матеріалу (адже це дослідження вищого академічного рівня!). Дослідження, які претендують на такий рівень фундаментальності, повинні ґрунтуватися на достовірному і переконливому цифровому матеріалі.

Виходячи з досвіду російських колег і попередніх результатів наших досліджень дисертації повинні стати об'єктом пильної уваги з боку АМН України, ВАК України і незалежних експертних груп, оскільки важко розраховувати на ефективність якої-небудь критики окремих ЖП по матеріалах КД. Це єдина реальна дорога поліпшення ситуації в області біостатистики і клінічної науки. Спочатку для такої роботи необхідно законодавчо забезпечити вільний доступ до авторефератів в мережі Інтернет як для захи-

щених дисертацій, так і за 3 місяці до їх захисту. Також слід забезпечити вільний доступ до тексту дисертацій. Виходячи з нашого досвіду, слід зауважити, що в багатьох установах (Національна бібліотека України ім. В.І. Вернадського, Дніпропетровська державна медична академія) до цих пір діють «циркуляри» і правила (на основі радянських законів), відповідно до яких доступ до тексту дисертацій вирішується лише по спеціальному листу, обмежений і утруднений здобуттям усіляких дозволів. Також забороняється необґрунтовано копіювати текст дисертацій, що в цілому не дозволяє провести якісну незалежну експертизу дисертаційного дослідження. У цьому питанні давно слід видати відповідний підзаконний акт на основі національного закону про інформацію. Другий аспект цієї експертної діяльності пов'язаний із створенням громадської організації і експертних груп, які зможуть забезпечити якісний і незалежний аудит поточної наукової продукції. Третій аспект вирішення проблеми пов'язаний з необхідністю створення простих рамок коректного представлення статистичного матеріалу КД у вигляді відповідного статистичного стандарту.

Літературні джерела

1. Барт А. Г. О состоянии дел и перспективах обучения статистическому анализу медицинских данных / А. Г. Барт, Е. В. Вербицкая, В. Н. Солнцев // *Международ. журн. мед. практи.* – 2006. – № 2. – С. 39–42.
2. Вараксин А. Н. Статистический анализ биологической и медицинской информации: проблемы и решения / А. Н. Вараксин // *Международ. журн. мед. практи.* – 2006. – № 2. – С. 35–38.
3. Власов В. В. Введение в доказательную медицину / Власов В. В. – М. : Медиа Сфера, 2001. – 392 с.
4. Власов В. В. Эпидемиология / Власов В. В. – М. : ГЭОТАР-МЕД, 2005. – 464 с.
5. Власов В. В. Учим и учимся статистике / В. В. Власов // *Международ. журн. мед. практи.* – 2006а. – № 2. – С. 5.
6. Власов В. В. Изучение методов диагностики / В. В. Власов // *Международ. журн. мед. практи.* – 2006б. – № 4. – С. 7–17.
7. Власов В. В. Систематические ошибки и вмешивающиеся факторы / В. В. Власов // *Международ. журн. мед. практи.* – 2007. – № 3. – С. 18–29.
8. Воробьев К. П. Проблемные вопросы представления материалов клинических исследований / К. П. Воробьев // *Біль, знеболювання і інтенсивна терапія.* – 2005а. – № 2. – С. 71–80.
9. Воробьев К. П. Оценка качества некоторых публикаций Европейского съезда анестезиологов 2005 года / К. П. Воробьев // *Біль, знебо-*

- лювання і інтенсивна терапія. – 2005б. – № 4. – С. 2–11.
10. Воробьев К. П. Формат современной журнальной публикации по результатам клинического исследования. Ч. 1. Сущность проблемы / К. П. Воробьев // *Укр. мед. часопис* – 2007. – Т. 6, № 62. – С. 18–26.
11. Воробьев К. П. Формат современной журнальной публикации по результатам клинического исследования. Ч. 2. Международные рекомендации / К. П. Воробьев // *Укр. мед. часопис* – 2008а. – Т. 1, № 63. – С. 58–66.
12. Воробьев К. П. Формат современной журнальной публикации по результатам клинического исследования. Ч. 3. Дизайн клинического исследования / К. П. Воробьев // *Укр. мед. часопис* – 2008б. – Т. 2, № 64. – С. 150–160.
13. Всемирная ассоциация медицинских редакторов / Рекомендации всемирной ассоциации медицинских редакторов по принципам этики научных публикаций в медицинских журналах // *Международ. журн. мед. практи.* – 2005. – № 5. – С. 24–30.
14. Гринхальх Т. Основы доказательной медицины / Гринхальх Т. – М. : ГЭОТАР-МЕД, 2004. – 240 с.
15. Дюк В. А. Преподавание прикладной статистики в Санкт-Петербургской медицинской академии последипломного образования / В. А. Дюк // *Международ. журн. мед. практи.* – 2006. – №

2. – С. 30–31.

16. Ланг Т. Двадцять ошибок статистического анализа, которые Вы сами можете обнаружить в биомедицинских статьях / Т. Ланг // *Международ. журн. мед. практи.* – 2005. – № 1. – С. 21–31.

17. Лапач С. Н. Статистика в науке и бизнесе / Лапач С. Н., Чубенко А. В., Бабич П. Н. – К. : МОРИОН, 2002. – 640 с.

18. Леонов В. П. Обучение медиков статистике: попытка системного подхода к проблеме / В. П. Леонов // *Международ. журн. мед. практи.* – 2006. – № 2. – С. 17–22.

19. Леонов В. П. Отечественная биостатистика: правильный диагноз — залог успешного лечения / В. П. Леонов // *Международ. журн. мед. практи.* – 2007. – № 1. – С. 78–80.

20. Петри А. Наглядная статистика в медицине / Петри А., Сэбин К. ; пер. с англ. – М. : ГОЭТАР-МЕД, 2003. – 144 с.

21. Постернак Г. І. Об'єктивізація шкал тяжкості і алгоритмів інтенсивної терапії у дітей з опіковим шоком (експериментально-клінічне дослідження) : автореф. дис. на здобуття наук.

ступеня д-ра мед. наук / Г. І. Постернак. - Дніпропетровськ, 2008. - 36 с.

22. Плавинский С. Л. О людях и цифрах. Обучение статистике: чему, кого и как учить? / С. Л. Плавинский // *Международ. журн. мед. практи.* – 2006. – № 2. – С. 9–16.

23. Реброва О. Ю. Открытое письмо главному редактору «Международного журнала медицинской практики» профессору В.В. Власову / О. Ю. Реброва // *Международ. журн. мед. практи.* – 2007. – № 1. – С. 71–72.

24. Тоффлер Э. Метаморфозы власти / Тоффлер Э. ; пер. с англ. – М. : АСТ, 2004. – 670 с.

25. Флетчер Р. Клиническая эпидемиология: основы доказательной медицины / Флетчер Р. , Флетчер С. , Вагнер Э. – М. : Медиа Сфера, 1998. – 350 с.

26. Хрипаченко І. А. Нейрогуморальна регуляція та оптимізація її оцінки у хворих на синдром поліорганної недостатності : автореф. дис. на здобуття наук. Ступеня д-ра мед. наук / І. А. Хрипаченко. - Дніпропетровськ, 2006. - 40 с.