

© Нечитайло Ю.М.

УДК 001.89

НАУКОВІ ДОСЛІДЖЕННЯ І БІОСТАТИСТИКА: І. "АНАТОМІЯ" ТА "ФІЗІОЛОГІЯ" НАУКОВИХ ДОСЛІДЖЕНЬ

Ю.М.Нечитайло

Кафедра пропедевтики дитячих хвороб (зав. – проф. Ю.М.Нечитайло) Буковинської державної медичної академії

Наука про методи збору, організації, обробки, аналізу та представлення даних, а також формування валідних висновків і прийняття результативних рішень на основі цього аналізу називається статистикою. Вона поділяється на математичну та прикладну. Частиною останньої є біостатистика, призначена для розв'язання медичних та біологічних проблем статистичними методами.

Статистика – це тільки інструмент. Медик часто, беручись за обробку своїх результатів без відповідного плану, схожий на столяра, який без кінця струже дошку, в надії, що з неї що-небудь вийде. Статистичні методи ефективні тільки у поєднанні із детальним плануванням та чіткою уявою про результати, які науковець хоче отримати.

Завдання дослідника: виконати наукову роботу швидко, дешево і просто. Але її результати мають бути доказовими і репрезентативними. Спробуємо допомогти молодому науковцю послідовно побудувати дослідження, підібрати відповідний статистичний інструмент і довести роботу до логічного завершення.

Поділимо наукове завдання на стадії виконання, які можуть бути внесені в список у логічній послідовності. При цьому кожна стадія залежить від попередньої.

1. Попередня стадія (актуальність дослідження). Роз'яснення мети, формулювання теми.
2. Планування.
3. Підготовка до збору даних.
4. Отримання даних.

5. Обробка результатів.

6. Отримання висновків.

7. Втілення результатів у життя.

Розрізняють побудову дослідження (анатомію) та виконання роботи (функціонування, фізіологію). Анатомія дослідження має такі складові:

- Основне питання, основна гіпотеза, мета і тема дослідження.
- Актуальність дослідження, його значення.
- Додаткові питання дослідження та завдання.
- Виконувальність роботи.
- Дизайн дослідження.
- Значимість та об'єм вибірки.
- Репрезентативність роботи. Спосіб формування вибірки, рандомізація.
- Об'єкт дослідження, випадок дослідження, змінна.
- Види даних, одиниці виміру і шкали. Сирі, стандартизовані і формалізовані дані. Сортування.
- Статистичні критерії, валідність.

Початком будь-якого дослідження є визначення напрямку та наукової і практичної цінності дослідження. Про його актуальність можна говорити, тільки проаналізувавши бібліографічні матеріали щонайменше останніх 5 років. Результатом попереднього етапу є чітка відповідь на запитання: навіщо виконується дана робота?

Наступним етапом є планування. Успіх, якість і швидкість виконання наукового дослідження залежать від його чіткої логічної побудови. Дослідник перш за все повинен

розуміти, що він робить, для чого і як. Загальний план дослідження подається у вигляді тематичної карти. Рекомендуємо для поточної роботи, ще більше його деталізувати, складаючи план на кожний етап і кожний крок дослідження.

Обов'язковими компонентами роботи є новизна та практична цінність. Якщо робота не вносить нічого нового, або не може бути застосованою у медичній чи науковій практиці – вона не є корисною і її не слід планувати.

Планування починається з формулювання основного питання і основної гіпотези дослідження. Будь-яка робота повинна базуватися на науковій ідеї, гіпотезі. Враховуючи невирішені проблеми, що впливають з вибраної гіпотези, формулюється основне питання, основна проблема дослідження. Основні проблема та ідея повинні формулюватися у вигляді одного чи двох коротких речень. Працюючи над матеріалом, бажано завжди мати перед собою назву вашої роботи, або основну її ідею. Одне, два речення копіюйте на початку аркуша з розрахунками, або під час написання розділів роботи. Це надасть вашій роботі логічної єдності.

Для прикладу, дослідження по виявленню артеріальної гіпертензії у молодих людей може бути сформульоване у вигляді основних питань дослідження – як часто трапляється артеріальна гіпертензія у молодих людей у віці до 20-30 років? Із основного питання формулюється мета дослідження та основна ідея. В наведеному прикладі основна ідея (гіпотеза) може бути сформована так: витоки артеріальної гіпертензії зрілого віку знаходяться у молодому віці.

Більш конкретні специфічні питання формують завдання дослідження.

В наведеному прикладі завдання можуть виглядати так:

- Залежність гіпертензії від віку молодої людини;
- Залежність від статі;
- Залежність від куріння;
- Залежність від вживання алкоголю;
- Генетична схильність.

Важливим є вирішення питання про фізичну можливість виконання роботи. Якщо патологія, що вивчається, є рідкісною і набрати необхідну кількість об'єктів дослід-

ження можливо тільки за 5-7 років, або якщо виконання роботи вимагає значних фінансових витрат чи великої кількості часу, то таке дослідження можна віднести до важко виконуваних чи нереальних у виконанні.

Поряд із загальним плануванням роботи, слід проводити поточне планування окремих фрагментів роботи. Поточне планування є складовою частиною протоколів виконання роботи. На початку протоколу бажано ретельно спланувати конкретні завдання виконуваного фрагменту роботи. В запис змісту роботи необхідно внести виконані дії та обов'язково відмітити, які результати отримані та в яких електронних файлах вони зберігаються.

Об'єкт дослідження. Окрема одиниця дослідження називається об'єктом дослідження. Це – матеріальний предмет, який має кількісну характеристику, а саме людина, експериментальна тварина, мікроорганізм, об'єднання людей (групи, установи) тощо. Матеріал дослідження піддається конкретному цифровому вираженню, тобто обстежено 8 хворих осіб, 25 груп студентів, 7 військових підрозділів, 10 білих щурів.

Предмет дослідження. Структура біологічного об'єкта, фізіологічний чи патологічний процес, професійні чи соціальні показники певної спільноти становлять предмет дослідження. Так, предметом дослідження може бути морфологічна структура м'язів ембріона (за гістологічним препаратом кінцівки), фізіологічні процеси в ендокринній системі, перебіг гепатиту під впливом лікування, соціальна думка сукупності людей з певного приводу тощо.

Певна множина об'єктів дослідження утворює вибірку. Результати дослідження пов'язані не з усією популяцією, а з обмеженою вибіркою, яка представляє цю популяцію. Формування вибірки повинно мати репрезентативний характер, що означає наявність у даної вибірки основних характеристик популяції в цілому. В основі досягнення репрезентативності лежить принцип випадковості, за яким будь-яка особа має шанси потрапити у вибірку для дослідження. Принцип випадковості має назву рандомізації (від англ. random – випадковий). Формування вибірки відбувається на основі випадкового використання

різних вторинних ознак, як то: день тижня, в який залучається особа до обстеження, номер історії чи картки обстеження (парний, непарний), порядковий номер студента у списку курсу, при телефонному опитуванні – кінцеві цифри номера телефону тощо.

В окремих випадках, коли досліджувана популяція має протилежні характеристики і вони є суттєвими для роботи, застосовується принцип стратифікації (від лат. strata – шар, прошарок). Цей підхід базується на розподілі всієї популяції на декілька прошарків та формуванні випадкової вибірки у кожному із прошарків на пропорційній основі. Для прикладу, дослідження артеріальної гіпертензії у студентів може проводитись за принципом статевої стратифікації. В цьому випадку окремо формується вибірка серед дівчат на випадковій основі і вибірка серед хлопців. Крім статевої, часто застосовують вікову стратифікацію. Враховуючи фізіологічні особливості новонароджених дітей, немовлят та дітей старше одного року, у більшості досліджень ці групи слід розрізняти. Суттєва різниця простежується і в старших вікових групах. Через це є нерациональними дослідження в групі віком 50-80 років. Таку групу слід розділити на більш однорідні вікові підгрупи.

Дизайн дослідження. Побудова роботи називається дизайном дослідження. Існує декілька стандартних програм побудови дослідження. У вітчизняній літературі та науковій практиці це поняття практично невідоме і не використовується. Розрізняють три основних види дизайну досліджень:

- Описовий (deskриптивний).
- Порівняльний (аналітичний).
- Експериментальний.

Залежно від тривалості та кількості обстежень дизайн буває одномоментним та динамічним, тривалим у часі (лонгітудинальним). В свою чергу лонгітудинальний дизайн може бути спрямованим у минуле (ретроспективний), чи у майбутнє (проспективний).

Найбільш поширеним є описовий (deskриптивний, від descriptio – опис) дизайн. Він може бути як самостійним, так і частиною двох інших видів дизайну. Deskриптивний дизайн не вимагає наявності декількох груп і може застосовуватись до всієї вибірки без її поділу на групи порівняння. В цьому виді

дослідження описуються якісні та кількісні характеристики об'єкта.

Аналітичний та експериментальний дизайн обов'язково будується на наявності щонайменше двох груп, які порівнюються – основної та контрольної. Аналітичний дизайн має три варіанти дослідження. Найпростішим є одномоментний (cross-sectional). Під одномоментним дослідженням розуміють одноразове обстеження, яке може тривати від декількох хвилин до декількох днів чи тижнів. Другим видом дослідження є когортне дослідження (cohort study). При цьому дизайні відбираються дві когорти – групи осіб, сформовані за окремими ознаками. Когортами дослідження можуть бути діти, народжені в одному пологовому будинку, робітники різних цехів тощо. Групи початково не повинні різнитися між собою за більшістю ознак, за винятком основного фактора впливу, що вивчається (наприклад, грудне чи штучне вигодовування; нормальні чи шкідливі виробничі умови). Третім є дизайн з побудовою за принципом „випадок-контроль” (case-control). Варіанти аналітичного дизайну розрізняються способом формування основної та контрольної груп.

Експериментальний дизайн поділяється на природний, лабораторний та клінічний. Побудова будь-якого клінічного дослідження з використанням фармакологічних препаратів відноситься саме до експериментального дизайну і повинна містити елементи подвійного сліпого контролю із застосуванням рандомізації та наявності плацебо (неактивний двійник фармакологічного препарату).

Медико-етичні норми. Сучасні вимоги до медико-біологічних досліджень включають обов'язкове дотримання етичних норм. Існують загальноприйняті протоколи клінічних, медико-біологічних та експериментальних досліджень з пацієнтами та з тваринами. Відповідно до цих протоколів не допускається застосування інвазійних методик без клінічних показань. Недопустимим є нанесення фізичної чи психологічної травми хворому. На будь-яке дослідження, не зумовлене його станом, необхідно отримати дозвіл пацієнта або його родичів. В експериментальних роботах регламентовано умови утримування тварин та методи забору біологічного ма-

теріалу. Без виконання етичних вимог міжнародні журнали не приймають до розгляду жодної наукової публікації.

Будь-який план виконання наукового дослідження перш за все вирішує питання об'єму конкретної вибірки. На цей об'єм впливають дві величини: з одного боку – фінансові та матеріальні витрати на дослідження, з другого – одержання достеменних результатів. Кінцевий результат повинен містити розкриті вірогідні причинно-наслідкові зв'язки.

Об'єкт дослідження поділяється на конкретні випадки (case). Іншим показником є кількість вимірювань, які проводяться у даного об'єкта. Кожне таке вимірювання, проведене у об'єкта дослідження, називають змінною – variable (рисунок).

До змінних відносяться так звані номінальні, які не мають виміру і описують об'єкт з якісної сторони, тобто вказують, що даний предмет відноситься до різних класів. Типовим прикладом змінної є стать, національність, місце проживання тощо. Другий вид змінних носить назву порядкових або рангових. В медико-біологічних дослідженнях вони використовуються досить часто. Формується певна шкала, за якою поділяється необхідна ознака. Для прикладу, куріння у студентів може бути класифіковано за ранговою шкалою на 3-4-5 підрозділів. Якщо студент не курить, ознака позначається як 0, якщо зрідка – 1, якщо часто, але невелику кількість (до 1 пачки) – 2, більше однієї пачки – 3. Шкала є довільною і може містити

декілька щаблів. Назву порядкової або рангової вона носить тому, що більш високий показник шкали вказує на більшу інтенсивність ознаки, але не має математичного змісту, тобто вдвічі більший показник не вказує на те, що ознака є вдвічі більшою.

Кількісна змінна, що містить цифрові дані, може бути безперервною чи інтервальною, але в будь-якому випадку показник можна обробити математично і вдвічі більший показник відповідає вдвічі інтенсивнішому виразу. До таких змінних відносяться лабораторні показники, показники артеріального тиску, температури та ін.

За відношенням мети дослідження змінні поділяються на залежні та незалежні. Незалежна змінна притаманна даному об'єкту і її вплив чи зв'язок з іншими показниками ми вивчаємо. Залежна змінна є тією результуючою величиною, яку ми бажаємо вивчити. Наприклад, стать людини, національність, місце проживання, куріння є незалежними змінними. Артеріальний тиск, частота серцевих скорочень можуть бути в дослідженні залежними змінними і відображати вплив профілактичних та лікувальних заходів. Правильний відбір залежних та незалежних змінних є необхідною умовою правильної побудови дослідження. Якщо в дослідженні не включена суттєва за впливом змінна, то всі результати втрачають вірогідність. Наприклад, якщо при вивченні факторів ризику гастроентерологічної патології не враховано режим та характер харчування – всі інші результати невірогідні.

	A	B	C	D	E	F
1	№	Name	Vik	stat	diagnoz	Glucosa
2		ПІП	Вік	Стать	Діагноз	Глюкоза
3			Кількісна змінна		Гастрит -1	
4				жін-0	Виразк.хвор.-2	Ммоль/л
5	<u>Випадок</u>			чол-1	Холецист - 3	
6					Гастрит-4	
7		1 Карповська А.А.	8	0	Номінальна змінна	4 5.7
8		2 Горбунов А.І.	5	1		4 5.3
9		3 Непомняща В.Г.	11	0		1 4.6
10		4 Зеленський І.І.	15	1		1 4.9
11		5 Федюшина В.О.	11	0		1 6.1
12		6 Константінов Є.О.	13	1		1 5.1
13		7 Бабич К.В.	12	0		1 4.9
14		8 Токмазова С.С.	11	0		1 6.2

Рис. Зразок оформлення електронної таблиці з даними.

Важливим практичним аспектом є запланований надлишок інформації. Початкова інформація має бути більш широкою та детальною із врахування всіх змін. У ході дослідження малоінформативні показники можуть бути виключені. Отримані дані поділяються на сирі та оброблені. Із початкових сирих даних у подальшому можна отримати додаткові показники шляхом групування (об'єднання декількох змінних), утворення похідних за певними формулами, переведення у відносні, ранжовані чи стандартизовані одиниці тощо.

Зберігання та захист інформації. Отримані результати з самого початку роботи повинні вноситись у комп'ютерні таблиці. Найбільш поширеними є такі електронні таблиці – Microsoft Excel (Microsoft Corp.), QuattroPro (Novell Inc.), StarOffice Calc. Вибір електронної таблиці залежить від наявності програми на комп'ютері. В перших рядках таблиці вносяться показники за назвою відповідних змінних. У наступних рядках під кожною назвою повинна йти розшифровка шкал та одиниць виміру показників (див. рис.). Якщо це рангова шкала, то вказується якому значенню шкали яке цифрове значення присвоєне. Якщо це лабораторно-інструментальні показники, то вноситься їх фізична розмірність (одиниці виміру – см, г/л тощо). Першу колонку залишаємо для нумерації, яка повинна бути єдиною для всіх випадків дослідження.

Сформований файл для зберігання даних бажано назвати скороченням (із 8 знаків), що вказує на його суть. Це краще робити латиницею із 6 букв та двох цифр. Наприклад: artysk01, що є скороченням повної назви – артеріальний тиск, файл перший. При подальшій модифікації внесення даних міняється номер цього файла з тим, щоб при помилкових змінах можна було повернутись до попереднього варіанта. При роботі на комп'ютерах із спільним доступом свою інформацію також слід захищати від сто-

ронніх осіб. Іноді справа доходить навіть до того, що дослідник може побачити статтю під чужим прізвиськом, написану на основі його багаторічного матеріалу. Сучасні комп'ютери дозволяють зберігати як робочі файли, так і кінцевий продукт у вигляді малюнків, схем, тез, чернеток статей тощо у шифрованому вигляді. Для цього слід вибрати функцію збереження файла під паролем.

Інформація у вигляді файлів з даними є найціннішим матеріалом наукового дослідження. Файли після кожного етапу роботи, крім збереження на комп'ютері, слід додатково зберігати на дискетах, дублюючи запис на двох окремих дискетах, одна з яких зберігається в надійному місці. Для збереження інформації слід брати нові дискети відомих фірм-виробників. Економія у декілька десятків копійок на купівлі більш дешевої дискети часто призводить до додаткових видатків на десятки чи сотні гривень. При наявності записувальних пристроїв на компакт-дисках слід використовувати і цей носій. Тривалість та якість зберігання інформації на компакт-дисках значно вища, ніж на магнітних носіях.

Окремою вимогою виконання роботи є ведення відповідної первинної документації, до якої відносяться журнали обліку, протоколи, виписки із історій хвороби, бібліографічні реферати, ксерокопії публікацій тощо. Для обліку роботи використовують стандартні журнали чи великі зошити. Всі сторінки в них мають бути пронумерованими, прошиті наскрізь нитками, кінці яких закриті наклейкою із печаткою відділу кадрів чи канцелярії. Саме в них ведеться облік усіх обстежень, дати забору матеріалу, номерів історій хвороби тощо. Якщо робота клінічного характеру, то всі результати повинні бути внесені чи вклеєні у історію хвороби, а виписка з неї відповідним чином завірена підписом (лікарняним статистиком, завідувачем відділенн чи начмедом) та печаткою лікувальної установи.

Література

1. Бейли Н. Математика в биологии и медицине. – М.: Мир, 1970. – 372 с.
2. Григорьев С.Г., Кувакин В.И., Николаевич М.С., Юнкеров В.И. Применение теории планирования эксперимента в медицинских исследованиях (включая персональный компьютер и пакет статистических программ Statistica for Windows): Учебное пособие / Под ред. В.И.Кувакина. – СПб., 1999. – 86 с.
3. Rosner B. Fundamental of biostatistics: 4th Edition. – NY: Duxbury Press, 1995. – 682 p.

Надійшла в редакцію 01.09.2003 р.