

МЕТОДИЧНИЙ ІНСТРУМЕНТАРІЙ

УДК: 159.99

Валерій Боснюк, кандидат психологічних наук, доцент

РОЛЬ ПОКАЗНИКА РОЗМІРУ ЕФЕКТУ В СУЧАСНИХ ПСИХОЛОГІЧНИХ ДОСЛІДЖЕННЯХ

Для підтвердження результатів дослідження в психологічних наукових роботах протягом багатьох років використовується процедура перевірки значущості нульової гіпотези (загальноприйнята абревіатура NHST – Null Hypothesis Significance Testing) із застосуванням спеціальних статистичних критеріїв. При цьому здебільшого значення статистики «р» (p-value) розглядається як еквівалент важливості отриманих результатів і сили наукових доказів на користь практичного й теоретичного ефекту дослідження. Таке некоректне використання та інтерпретації p-value ставить під сумнів застосування статистики взагалі та загрожує розвитку психології як науки. Ототожнення статистичного висновку з науковим висновком, орієнтація виключно на новизну в наукових дослідженнях, ритуальна прихильність дослідників до рівня значущості 0,05, опора на статистичну категоричність «так/ні» під час прийняття рішення призводить до того, що психологія примножує тільки результати про наявність ефекту без врахування його величини, практичної цінності.

Дана робота призначена для аналізу обмеженості p-value при інтерпретації результатів психологічних досліджень та переваг представлення інформації про розмір ефекту. Застосування розмірів ефекту дозволить здійснити перехід від дихотомічного мислення до оціночного, визначати цінність результатів незалежно від рівня статистичної значущості, приймати рішення більш раціонально та обґрунтовано.

Обґрунтовується позиція, що автор наукової роботи при формулюванні висновків дослідження не повинен обмежуватися одним єдиним показником рівня статистичної значущості. Осмислені висновки повинні базуватися на розумному балансуванні p-value та інших не менш важливих параметрів, одним з яких виступає розмір ефекту. Ефект (відмінність, зв'язок, асоціація) може бути статистично значущим, а його практична (клінічна) цінність – незначною, тривіальною. «Статистично значущий» не означає «корисний», «важливий», «цінний», «значний». Тому звернення уваги психологів до питання аналізу виявленого розміру ефекту має стати обов'язковим при інтерпретації результатів дослідження.

Ключові слова: розмір ефекту, оцінка розміру ефекту, процедура перевірки значущості нульової гіпотези, p-value.

Вступ. Суть проблеми p-value полягає в тому, що не всі дослідники розуміють зміст процедури перевірки значущості нульової гіпотези, і як результат, некоректно інтерпретують результати.

Найбільш розповсюдженими помилками є інтерпретація значення статистичної значущості як 1) показника істинності

нуль-гіпотези, 2) міри ймовірності випадкового отримання результатів, 3) об'єктивного засобу перевірки нуль-гіпотези, 4) міри доказу істинності альтернативної гіпотези, 5) міри відтворення результатів та 6) показника практичної важливості отриманих результатів [6]. У дійсності, p-value – це розрахована ймовірність

того, що даний або більш екстремальний результат для вибірки визначеного обсягу може бути отриманий за умови, що нуль-гіпотеза вірна в генеральній сукупності [15]. Тобто, значення статистичної значущості нам показує ймовірність отримати нашу вибірку, якщо вірна H_0 , а нас цікавить ймовірність H_0 , якщо дана наша вибірка, що не одне й те ж саме.

Значення p-value залежать не тільки від отриманих даних, а й від чинників, які перебувають під контролем дослідника: обсягу вибірки, обраного статистичного критерію, дизайну дослідження, формулювання нуль-гіпотези. Так, на підставі суб'єктивних уявлень дослідник може ділити p-value навпіл, щоби використовувати односторонній статистичний критерій або коригувати його в процедурі множинних порівнянь. Особлива необ'єктивність проявляється в його чутливості до обсягу вибірки. Ми можемо отримати будь-яке «необхідне» низьке значення статистичної значущості збільшуючи обсяг вибірки.

Зростання невдоволення p-value призвело до появи нових мір визначення значущості результатів дослідження, таких як D-value [14] та S-value [17].

Дж. Коен заявив, що «NHST не тільки не сприяла прогресу психології як науки, але й серйозно перешкоджала цьому» [13, с. 997]. На думку Р. Кірка «наука заплатила високу ціну за свою ритуальну прихильність до перевірки нуль-гіпотези» [21, с. 756]. Акцент на значенні p-value і відхиленні нульової гіпотези відволікає нас від дійсних дослідницьких цілей: прийняття рішення, чи підкріплюють емпіричні

дані нашу наукову гіпотезу й чи є вони практично цінними та корисними [21].

З метою підвищення якості наукових робіт та переоцінки статистичної практики прийняття рішень було здійснено низку заходів, серед яких найбільш простим та логічним є інформування в дослідженнях про виявлений розмір ефекту.

Аналіз останніх досліджень та публікацій. У 1994 році Американська психологічна асоціація (АРА) вперше повідомила авторів, що у своїх дослідженнях їм «рекомендується» звітувати про розмір ефекту [8, с. 18]. Проте таке формулювання, особливо на фоні всіх інших суворих вимог фактично нічого не змінило в практиці застосування розміру ефекту.

Пізніше, у 1999 році Л. Вілкінсон та спеціальна робоча група АРА зі статистичних висновків підготували звіт, у якому зазначалося: «Дослідники повинні завжди представляти розмір ефекту для основних результатів» [28, с. 599]. Дана заява була реакцією на попередню м'яку вимогу зазначення розміру ефекту в наукових роботах. Вони роз'яснили свою позицію: «Ми наголошуємо, що повідомлення та інтерпретація розмірів ефекту в контексті раніше отриманих результатів є невід'ємною частиною якісного дослідження. Це дає можливість читачеві оцінити стабільність результатів щодо вибірок, дизайнів і способів аналізу даних. Інформація про розміри ефекту необхідна для аналізу потужності і здійснення метааналізу в подальших дослідженнях» [28, с. 599].

Вищезазначене відобразилось уже у 5-му виданні 2001 року «Керівництва Американської психологічної асоціації з підготовки

публікацій» [9], що спонукало редакторів наукових журналів вимагати не тільки описову статистику та результати перевірки значущості нульової гіпотези, але й розмір ефекту.

У лютому 2015 року редактори достатньо відомого журналу «Basic and Applied Social Psychology» (BASP) взагалі перестали приймати до опублікування статті, у яких використовувались процедури перевірки значущості нульової гіпотези та довірчих інтервалів [25]. У редакційній статті вказується, що значення «р» стало опорою для вчених, які працюють зі слабкими даними. Вони пишуть, що рівень $p < 0,05$ дуже легко обійти та іноді є виправданням для досліджень низької якості. Було зазначено, що перед публікацією в журналі авторам у своїх статтях доведеться видалити всі елементи NHST (р-value, t-значення, F-значення, твердження про статистичну значущість відмінностей або взаємозв'язку тощо). У заміну редактори вимагали суворої описової статистики із зазначенням розмірів ефекту. Заохочувалося подання відомостей про частоти або розподіли даних. Було звернено увагу, що психологи здебільшого формують недостатні за обсягом вибірки, а збільшення кількості досліджуваних дозволить зробити описову статистику більш стабільною. Водночас, редактори журналу залишили за собою право приймати індивідуальні рішення про наукові роботи, що базуються на баєсовому підході в статистиці. Були опубліковані конкретні рекомендації дослідникам щодо організації та представлення результатів своїх досліджень [26]. Ці вимоги склалися з трьох основних принципів аналізу даних: 1) використання графічних форм представлення результатів із

метою полегшення розуміння описової статистики; 2) надання інформації про міри мінливості даних поряд із мірами центральної тенденції; 3) розрахунок та коректна інтерпретація розмірів ефекту.

У березні 2016 року для зупинки зловживань в перевірці статистичної значущості Американська статистична асоціація (ASA) опублікувала заяву за результатами обговорень та роботи спеціальної комісії, що стосується використання р-value в наукових дослідженнях. Згідно з їх висновків, зазвичай р-value неправильно інтерпретується, а процедуру перевірки значущості нульової гіпотези неправильно застосовують, що призводить до «розуміння», яке в більшості випадків є безглуздою випадковістю. У заключних зауваженнях було, зокрема, відзначено, що р-value не дає достатніх доказів щодо підтвердження моделі або гіпотези та не є показником ефекту [27].

Після оголошення про заборону NHST та опублікованої заяви ASA Р. Меттьюз, Р. Вассерстейн, Д. Шпігельхальтер проаналізували, що змінилося за рік із процедурою перевірки значущості нульової гіпотези [22]. Автори визнали, серед статистиків відсутній консенсус, який підхід до аналізу даних кращий в епоху після $p < 0,05$. Вони зазначають, що це й не дивно та не варто очікувати єдиного фундаментального підходу для вирішення будь-якої логічної проблеми. Однак відсутність консенсусу не означає, що немає ефективного рішення. Проблема полягає не стільки в самому р-value, скільки в готовності дослідників формулювати впевнені узагальнюючі висновки з недбало описаних даних, отриманих у погано спланованих

дослідженнях [22, с. 41]. У кінцевому рахунку, кожен науковець у своїй галузі зобов'язаний дотримуватися етичних вимог до організації дослідження та аналізу даних. Усе, що може зробити товариство статистиків, – це надати інструменти аналізу, а дослідник повинен правильно застосовувати їх та належним чином розшифровувати результати (інтерпретувати).

Рональд Д. Фрикер, К. Берк, С. Хан, Уільям Х. Вудолл проаналізували наукові роботи опубліковані у журналі *BASP* після заборони використання процедури перевірки значущості нульової гіпотези та довірчих інтервалів. На їхню думку, це обмеження не змінило ситуацію на краще [16]. Було відзначено, що за відсутності міри статистичної значущості автори статей зосереджуються тільки на показниках розміру ефекту, що призводить до потенційного завищення результатів дослідження. Вони вважають, що розмір ефекту все ж таки необхідно зазначати разом з *p-value*. На думку авторів статті, проблема статистичного висновку полягає в тому, що дослідники психологічних явищ заціклені тільки на доказі своїх гіпотез, вони не сфокусовані на питанні чи можна пояснити виявлений ефект випадковою мінливістю даних. У цьому контексті *p-value* – це перша лінія захисту від випадкового обману, що відділяє шум у даних від реального ефекту.

Слід відмітити, що в даній ситуації Американська психологічна асоціація не відмовилася від своїх раніше задекларованих у 6-му виданні вимог до наукових публікацій наголошуючи, що перевірка значущості нульової гіпотези є лише початковим етапом проведення

наукового дослідження [10]. «Мінімальні очікування для всіх журналів АРА – повна звітність про всі перевірені гіпотези, оцінка відповідних розмірів ефектів, представлення описової статистики та довірчих інтервалів» [10, с. 33]. Дана позиція відобразилась і в сучасному 7-му виданні вимог до наукових публікацій 2020 року [11].

Щодо інших наукових організацій, то Американська асоціація освітніх досліджень (AERA) з 2006 року рекомендує, щоб статистичні результати досліджень включали різноманітні типи розміру ефекту, а також відповідні довірчі інтервали [7]. Міжнародний комітет редакторів медичних журналів (ICMJE) у рекомендаціях 2014 року щодо проведення, звітності, редагування та публікації наукових робіт зазначив: «При прийнятті рішення не покладайтеся лише на перевірку статистичних гіпотез, *p-value* не в змозі передати інформацію про розмір ефекту та точність оцінок» [19, с. 14].

Проте, незважаючи на рекомендації та сучасні тенденції у світовій психології у вітчизняних наукових журналах редактори вимагають, а дослідники у своїх роботах звітують лише про описову статистику та результат перевірки значущості нульової гіпотези. Читач наукової роботи повинен знати не тільки про те, що середні арифметичні значення в групах А і Б статистично значуще відрізняються, але й цінність цієї різниці, тобто розмір ефекту.

З огляду на зазначене, **мета дослідження** полягає у аналізі переваг представлення в наукових працях окрім показника статистичної значущості й інформацію про розмір виявленого ефекту.

Виклад основного матеріалу.

Проблему вимірювання розміру ефекту актуалізував Джейкоб Коен, як частини розробленої ним парадигми метааналізу [12]. Він зміг донести сенс та важливість поняття «розмір ефекту» до спеціалістів соціальних наук та розробив найбільш популярні індекси оцінки її величини. Позначається аббревіатурою ES (effect size). В деяких роботах можна зустріти вирази: «величина статистичного ефекту», «величина стандартизованого ефекту» [2; 3; 6].

Розмір ефекту – кількісне відображення величини (ступеня прояву) деякого явища, що представляє науковий інтерес [20, с. 140]. Він відображає силу взаємозв'язку між явищами, величину різниці ознаки між групами, відповідність теоретичної моделі емпіричним даним та дозволяє оцінити корисність, важливість, цінність, осмисленість такого відношення. Розмір ефекту за своєю суттю близький до поняття практичної значущості дослідження – головного орієнтира наукової роботи.

Розміри ефекту дозволяють дослідникам вийти за межі дихотомічного мислення: «Є відмінність (взаємозв'язок) чи немає?» до більш важливого аспекту: «Яка величина цих відмінностей (сила взаємозв'язку)?», відійти від акцентуванні на ідентифікації статистичної значущості та перейти до більш цінного і простішого для розуміння кількісного опису величини розміру ефекту, більш наукового підходу до накопичення знань. Окрім того класичні показники розміру ефекту не залежить від обсягу вибірки на відміну від статистичної значущості, яку за це постійно критикують.

У якості показника розміру ефекту можуть виступати нестандартизовані величини (різниця між двома середніми арифметичними значеннями, відсотками, нестандартизовані регресійні коефіцієнти тощо) та стандартизовані індекси, які розраховуються для представлення величини ефекту в зрозумілому масштабі.

Нестандартизовані величини розміру ефекту корисні, коли змінні дослідження мають безпосередньо зрозуміле значення (наприклад, швидкість прийняття рішення). Стандартизовані індекси величини ефекту необхідно застосовувати, коли вимірювання не має внутрішнього значення, наприклад, числа в шкалі Лайкерта; коли в дослідженнях використовувалися різні шкали, тому пряме порівняння результатів неможливе; або коли розмір ефекту розглядається в контексті мінливості в популяції.

Вибір індексу розміру ефекту (див. табл.1) в конкретному дослідженні залежить від шкали вимірювання, використаних статистичних критеріїв та дизайну дослідження [більш детально 1; 18].

Необхідно зазначити, що наведені стандартизовані індекси ES не єдино можливі для зазначених статистичних критеріїв. Звернувшись до спеціалізованої літератури можна виявити альтернативні міри оцінки розміру ефекту. Крім того, вони майже всі пов'язані між собою, за необхідності можна трансформувати одні індекси в інші. Така універсальність дозволяє порівнювати результати різних досліджень у метааналітичних дослідженнях для отримання більш надійних оцінок параметрів генеральної сукупності.

Табл.1. Популярні індекси кількісної оцінки розміру ефекту та їх рекомендована інтерпретація

Статистичний критерій	Індекс ES	Розмір ефекту			
		несуттєвий	малий	середній	великий
Тест Стьюдента	d	0,00 - 0,19	0,20 - 0,49	0,50 - 0,79	0,80 >
Манна-Уїтні, Вілкоксона	r	0,00 - 0,09	0,10 - 0,29	0,30 - 0,49	0,50 - 1,00
Дисперсійний аналіз	η^2	0,00 - 0,009	0,01 - 0,05	0,06 - 0,13	0,14 - 1,00
Регресійний аналіз	f^2	0,00 - 0,01	0,02 - 0,14	0,15 - 0,34	0,35 - 1,00
Кореляційний аналіз	r	0,00 - 0,09	0,10 - 0,29	0,30 - 0,49	0,50 - 1,00
χ^2 -Пірсона	w	0,00 - 0,09	0,10 - 0,29	0,30 - 0,49	0,50 - 1,00

Оскільки стандартизовані величини ефекту не мають одиниць вимірювань, їхня інтерпретація для кожного індексу зводиться до питання який ефект вважається малим, середнім або великим. Традиційно для вирішення даної проблеми застосовується два підходи. Перший із них є порівняння виявленого ефекту з типовими ефектами отриманими іншими дослідниками у відповідній галузі знань. Другим підходом при інтерпретації є орієнтація на «еталонні» значення. Найбільш визнані в поведінкових науках є методичні рекомендації Дж. Коена (див. табл.1).

Тим не більше, жоден із традиційних підходів до інтерпретації розміру ефекту в психологічній галузі знань не позбавлений серйозних проблем [24]. Порівняння виявленого ефекту з іншими дослідниками ускладнюється низькою відтвореністю результатів експериментів. Наприклад, у статті над якою працювали 270 осіб був представлений звіт про повторне відтворення 100 досліджень, що були опубліковані в провідних психологічних журналах. Як

результат, понад 60 % реплікацій не підтвердили статистичну значущість висновків на рівні 0,05. Середній розмір ефекту (r) реплікацій ($M_r = 0,197$, $SD = 0,257$) був удвічі менший у порівнянні з опублікованими результатами ($M_r = 0,403$, $SD = 0,188$), що відображає суттєве зменшення [23].

З іншої сторони, використання еталонних стандартів у психології під час інтерпретації величини розміру ефекту втрачає сенс через значну різномірність психологічних наукових напрямів. Т. Шефер та М. Шварц [24] виявили значну відмінність у медіанах розподілів розмірів ефекту в різних галузях психології. Найбільші показники розмірів ефекту публікуються в дослідженнях з експериментальної та біологічної психології, а в соціальній психології та психології розвитку – значно менші.

Одне з прагматичних рішень для розв'язання проблеми інтерпретації запропонував сам Дж. Коен [12]: відображати розмір ефекту в нестандартизованих величинах та інтерпретувати їхнє практичне значення в контексті проведеного дослідження.

Тобто запропоновані узагальнені градації розмірів ефекту необхідно сприймати тільки в якості орієнтирів, бо для кожного окремого дослідження повинен застосовуватися індивідуальний підхід. Оцінюючи необхідно враховувати теоретичні, економічні, етичні та практичні аспекти.

Обговорення результатів. Існує декілька типових ситуацій інтерпретації результатів дослідження з урахуванням оцінки розміру ефекту після тестування гіпотези:

1) якщо немає достатніх доказів для відхилення H_0 (не виявили відмінностей), а ES середній чи великий, не можна стверджувати про відсутність відмінностей. Це привід здійснити додаткове дослідження збільшивши обсяг вибірки;

2) якщо H_0 не відхилено, а ES несуттєвий чи малий, можна констатувати, що відмінностей швидше всього дійсно немає;

3) якщо є підстави для відхилення H_0 (виявили відмінності), а ES несуттєвий чи малий, необхідно добре подумати чи мають такі відмінності психологічне значення (можливо вони взагалі в межах помилки вимірювання);

4) якщо H_0 відхилено, а ES великий є підстави стверджувати, що відмінності дійсно існують.

Оцінка розміру ефекту до тестування гіпотези особливо актуальна в експериментальних дослідженнях. Вона надає можливість розрахувати необхідний обсяг вибірки під час планування роботи, що дуже важливо при обмеженому доступі до досліджуваних. У такому випадку, необхідно додатково визначити прийнятний рівень помилку першого роду та потужність дослідження. Такі розрахунки виконуються, наприклад, у

калькуляторі вільного доступу G*Power або програмно-статистичним середовищем із відкритим кодом R.

Окрім того, увага до розміру ефекту не вичерпується цікавістю до конкретного дослідження, важливим аспектом цього інтересу є бажання уніфікувати звіти про результати одиничних досліджень так, щоби їх можна було використовувати в метааналітичних проєктах. З цією метою науковому товариству необхідно звернути увагу на стандарти розроблені APA з публікації емпіричних даних, тому що раніше висловлене твердження С.В. Сивухою не втратило своєї актуальності й сьогодні: «У нашому розпорядженні немає вихідного матеріалу для метааналізу – достатньої кількості публікацій, які присвячені одній темі та написані відповідно до норм співставлення: детальним і чітким описом процедурних особливостей дослідження та його результатів в термінах статистичних ефектів. У вітчизняній психології немає репрезентативною бази даних публікацій, що містять релевантну статистичну інформацію» [5, с. 38].

Отже, є достатньо вагомими причини дослідникам доповнити свої звіти про перевірку значущості нульової гіпотези інформацією про розмір ефекту та його довірчі інтервали. У якості основних аргументів виділяють наступні:

1. На відміну від значення p-value розмір ефекту відображає цінність результатів дослідження.

2. Дозволяє оцінити ступінь прояву явища незалежно від обсягу вибірки.

3. Представлені результати дослідження в стандартизованому вигляді дозволяють порівнювати між

собою ознаки, що діагностовано в різних вимірjuвальних шкалах.

4. Використовується для визначення необхідного обсягу вибірки в запланованих дослідженнях.

5. Дозволяє оцінити статистичну потужності у вже проведених дослідженнях.

6. Розмір ефекту незамінний у проведенні метааналітичних досліджень, дозволяючи порівнювати між собою результати різних авторів.

7. Представлений разом із довірчим інтервалом розмір ефекту може використовуватися для оцінювання параметрів генеральної сукупності.

Висновки. Для більшої об'єктивності при інтерпретації результатів дослідження важливо в наукових звітах крім аналізу статистичної значущості представити дані про величину розміру ефекту і його довірчі інтервали. Не дивлячись на безумовну цінність цієї інформації та рекомендації міжнародних психологічних асоціацій, очевидно, що у вітчизняних дослідженнях не практикується її використання. Необхідно заохочувати популяризацію даної методології, включити її в освітні програми з підготовки психологів. Вимога представлення розмірів ефектів у науковому дослідженні стане відповідати і змістовному завданню науки – можливість статистичного узагальнення різних емпіричних досліджень, оцінка стійкості результатів.

Перспективи подальших досліджень. Через низьку поінформованість науковців існує необхідність проаналізувати, навести приклади розрахунків та інтерпретації різноманітних індексів розмірів ефекту та їхніх довірчих інтервалів

залежно від використаних статистичних критеріїв у психологічних дослідженнях..

Література

1. Боснюк В. Ф. Математичні методи в психології: курс лекцій. Мультимедійне навчальне видання. Х. : НУЦЗУ, 2020. 141 с.

2. Корнеев А. А., Рассказова Е. И., Кричевец А. Н., Койфман А. Я. Критика методологии проверки нулевой гипотезы: ограничения и возможные пути выхода. Часть I. *Психологические исследования*. 2016. Т. 9. № 45. Режим доступа: <http://psystudy.ru/index.php/num/2016v9n45/1231-korneev45.html#r1>

3. Кричевец А. Н., Корнеев А. А., Рассказова Е. И. Основы статистики для психологов. Москва : Акрополь, 2019. 286 с.

4. Куприянов И. В. Расчёт величины эффекта при применении непараметрических критериев сравнения выборок в психологических исследованиях с использованием программы R. *Электронный научный журнал «Инженерный вестник Дона»*. 2016. № 2. Режим доступа: ivdon.ru/ru/magazine/archive/n2y2016/3610

5. Сивуха С. В. От новизны результатов к воспроизводимости: новые ориентации в психологии. *Вестні БДПУ*. 2012. Серія 1. № 2. С. 35–39.

6. Сивуха С. В., Козяк А. А. О реформе статистического вывода в психологии. *Психология. Журнал Высшей школы экономики*. 2009. Т. 6. № 4. С. 66–86.

7. American Educational Research Association. Standards for Reporting on Empirical Social Science Research. *Educational Researcher*. 2006. Vol. 35. Issue 6. P. 33–40.

<https://doi.org/10.3102/0013189X035006033>

8. American Psychological Association. Publication Manual of the American Psychological Association : 4-th ed., Washington, DC : Author, 1994.

9. American Psychological Association. Publication Manual of the American Psychological Association : 5-th ed., Washington, DC : Author, 2001. 439 p.

10. American Psychological Association. Publication Manual of the American Psychological Association : 6-th ed., Waschinton, DC : Author, 2010. 272 p.

11. American Psychological Association. Publication manual of the American Psychological Association : 7-th ed., Waschinton, DC : Author, 2020. 428 p.

12. Cohen J. Statistical Power Analysis for the Behavioral Sciences. 2nd Edn. New York : Lawrence Earlbaum Associates, 1988. 567 p.

13. Cohen J. The earth is round ($p < .05$): Rejoinder. *American Psychologist*. 1994. Vol. 50. Issue 12. P. 997–1003.
<https://doi.org/10.1037%2F0003-066x.49.12.997>

14. Demidenko E. The p-Value You Can't Buy. *The American Statistician*. 2016. Vol. 70. Issue 1. P. 33–38.
<https://doi.org/10.1080/00031305.2015.1069760>

15. Finch S., Cumming G., Thomason N. Reporting of statistical inference in the Journal of Applied Psychology: Little evidence of reform. *Educational and Psychological Measurement*. 2001. Vol. 61. Issue 2. P. 181–210.
<https://doi.org/10.1177/00131640121971167>

16. Fricker R., Burke K., Han X., Woodall W. Assessing the Statistical Analyses Used in Basic and Applied Social Psychology After Their p-Value Ban. *The American Statistician*. 2019. Vol. 73. P. 374–384.
<https://doi.org/10.1080/00031305.2018.1537892>

17. Greenland S. Valid P-Values Behave Exactly as They Should: Some Misleading Criticisms of P-Values and Their Resolution With S-Values. *The American Statistician*. 2019. Vol. 73. P. 106–114.
<https://doi.org/10.1080/00031305.2018.1529625>

18. Ialongo C. Understanding the effect size and its measures. *Biochemia medica*. 2016. Vol. 26. Issue 2. P. 150–163.
<https://doi.org/10.11613/BM.2016.015>

19. International Committee of Medical Journal Editors. Recommendations for the conduct, reporting, editing, and publication of scholarly work in medical journals. 2014. Retrieved from www.icmje.org.

20. Kelley K., Preacher K. J. On effect size. *Psychological Methods*. 2012. Vol. 17. Issue 2. P. 137–152.
<https://doi.org/10.1037/a0028086>

21. Kirk R. E. Practical significance: A concept whose time has come. *Educational and Psychological Measurement*. 1996. Vol. 56. Issue 5. P. 746–759.
<https://doi.org/10.1177/0013164496056005002>

22. Matthews R., Wasserstein R. L., Spiegelhalter D. The ASA's p-value statement, one year on. *Significance*. 2017. Vol. 14. P. 38–41.
<https://doi.org/10.1111/j.1740-9713.2017.01021.x>

23. Open Science Collaboration. Estimating the reproducibility of psychological science. *Science*. 2015.

- Vol. 349. Issue 6251. aac4716. P. 260–
<https://doi.org/10.1126/science.aac4716> 273. doi:10.1080/01973533.2015.10602
24. Schäfer T., Schwarz M. A. The 40
Meaningfulness of Effect Sizes in
Psychological Research: Differences
Between Sub-Disciplines and the Impact
of Potential Biases. *Frontiers in
psychology*. 2019. Vol. 10. Article 813.
<https://doi.org/10.3389/fpsyg.2019.00813>
3
25. Trafimow D., Marks M.
Editorial. *Basic and Applied Social
Psychology*. 2015. Vol. 37. P. 1–2.
<https://doi.org/10.1080/01973533.2015.1012991>
26. Valentine J. C., Aloe A. M.,
Lau T.S. Life After NHST: How to
Describe Your Data Without “p-ing”
Everywhere. *Basic and Applied Social
Psychology*. 2015. Vol. 37. Issue 5.
27. Wasserstein R. L., Lazar N. A.
The ASA Statement on p-Values:
Context, Process, and Purpose. *The
American Statistician*. 2016. Vol. 70.
Issue 2, P. 129–133.
<https://doi.org/10.1080/00031305.2016.1154108>
28. Wilkinson L., Task Force on
Statistical Inference, American
Psychological Association, Science
Directorate. Statistical methods in
psychology journals: Guidelines and
explanations. *American Psychologist*.
1999. Vol. 54. Issue 8. P. 594–
604. <https://doi.org/10.1037/0003-066X.54.8.594>

Valerii Bosniuk, Ph.D. in Psychology, Associate Professor

THE ROLE OF THE EFFECT SIZE INDICATOR IN MODERN PSYCHOLOGICAL RESEARCH

The procedure for testing the significance of the null hypothesis (common abbreviation NHST - Null Hypothesis Significance Testing) is have being used for confirmation of research results in psychological research for many years and it also includes special statistical criteria. At the same time, for the most part, the value of Statistics «p» (p-value) is considered as the equivalent of the importance of the results obtained and the strength of scientific evidence in favor of the practical and theoretical effect of the study. This incorrect use and interpretation of p-value impugns the application of Statistics in general and threatens the development of psychology as a science. The identification of a statistical conclusion with a scientific conclusion, the focus solely on novelty in scientific research, the ritual commitment of researchers to the significance level of 0.05, the reliance on the statistical categorical "yes/no" when making a decision leads to the fact that psychology multiplies only the results about the presence of an effect without taking into account its magnitude, practical value.

This work is intended to analyze the limitations of p-value when interpreting the results of psychological research and the benefits of presenting information about the effect size. Applying the size of the effect will allow you to make the transition from dichotomous thinking to evaluative thinking, determine the value of results regardless of the level of statistical significance, and make decisions more rationally and reasonably.

The position that the author of the scientific work in formulating the study conclusions should not be limited to a single indicator of the level of statistical significance is substantiated. Reasonable conclusions should be based on a reasonable balancing of p-value and other equally

important parameters, one of which is the size effect. The effect (difference, connection, association) can be statistically significant, and its practical (clinical) value – insignificant, trivial. «Statistically significant» does not mean «useful», «important», «valuable», «considerable». Therefore, the attention of psychologists to the analysis of the detected size effect should be mandatory when interpreting the results of the study.

Keywords: effect size, effect size estimation, null hypothesis significance testing, p-value.

References

1. Bosniuk, V. F. (2020). *Matematychni metody v psykholohii: kurs lektsii. Multymediine navchalne vydannia [Mathematical methods in psychology: a course of lectures. Multimedia educational publication]*. Kh.: NUTsZU. [in Ukrainian].
2. Korneev, A. A., Rasskazova, E. I., Krichevec, A. N., & Kojfman, A. YA. (2016). Kritika metodologii proverki nulevoj gipotezy: ogranicheniya i vozmozhnye puti vyhoda. CHast' I [Criticism of Null Hypothesis Significance Testing: Limitations and Possible Ways Out. Part I]. *Psihologicheskie issledovaniya*, 9(45), 1. Retrieved from <http://psystudy.ru/index.php/num/2016v9n45/1231-korneev45.html#r1> [in Russian].
3. Krichevec, A. N., Korneev, A. A., & Rasskazova, E. I. (2019). *Osnovy statistiki dlya psihologov [Fundamentals of statistics for psychologists]*. Moskva: Akropol'. [in Russian].
4. Kupriyanov, I. V. (2016). Raschyot velichiny efekta pri primenenii neparametricheskikh kriteriev sravneniya vyborok v psihologicheskikh issledovaniyah s ispol'zovaniem programmy R [Calculation of the effect sizes when applying nonparametric criteria for comparing samples in psychological research using the R program]. *Elektronnyj nauchnyj zhurnal «Inzhenernyj vestnik Dona»*, 2. Retrieved from ivdon.ru/ru/magazine/archive/n2y2016/3610 [in Russian].
5. Sivuha, S. V. (2012). Ot novizny rezul'tatov k vosproizvodimosti: novye orientacii v psihologii [From the novelty of results to reproducibility: new orientations in psychology]. *Vesci BDPU*, 2(1), 35–39. [in Russian].
6. Sivuha, S. V., & Kozyak, A. A. (2009). O reforme statisticheskogo vyvoda v psihologii. Psihologiya [On the reform of statistical inference in psychology]. *ZHurnal Vysshej shkoly ekonomiki*, 6(4), 66–86. [in Russian].
7. American Educational Research Association. (2006). Standards for Reporting on Empirical Social Science Research in AERA publications. *Educational Researcher*, 35(6), 33–40. <https://doi.org/10.3102/0013189X035006033>
8. American Psychological Association. (1994). *Publication Manual of the American Psychological Association* (4th ed.). Washington, DC: Author.
9. American Psychological Association. (2001). *Publication Manual of the American Psychological Association* (5th ed.). Washington, DC: Author.
10. American Psychological Association. (2010). *Publication Manual of the American Psychological Association* (6th ed.). Waschinton, DC: Author.
11. American Psychological Association. (2020). *Publication manual of the American Psychological Association* (7th ed.). American Psychological Association. <https://doi.org/10.1037/0000165-000>

12. Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd edition). New York : Lawrence Earlbaum Associates.
13. Cohen, J. (1994). The earth is round ($p < .05$): Rejoinder. *American Psychologist*, 50(12), 997–1003. <https://doi.org/10.1037%2F0003-066x.49.12.997>
14. Demidenko, E. (2016). The p -Value You Can't Buy. *The American statistician*, 70(1), 33–38. <https://doi.org/10.1080/00031305.2015.1069760>
15. Finch, S., Cumming, G., & Thomason, N. (2001). Reporting of statistical inference in the Journal of Applied Psychology: Little evidence of reform. *Educational and Psychological Measurement*, 61(2), 181–210. <https://doi.org/10.1177/00131640121971167>
16. Fricker, R., Burke, K., Han, X., & Woodall, W. (2019). Assessing the Statistical Analyses Used in Basic and Applied Social Psychology After Their p -Value Ban. *The American Statistician*, 73, 374–384. <https://doi.org/10.1080/00031305.2018.1537892>
17. Greenland, S. (2019). Valid P -Values Behave Exactly as They Should: Some Misleading Criticisms of P -Values and Their Resolution With S -Values. *The American Statistician*, 73, 106–114. <https://doi.org/10.1080/00031305.2018.1529625>
18. Ialongo, C. (2016). Understanding the effect size and its measures. *Biochemia medica*, 26(2), 150–163. <https://doi.org/10.11613/BM.2016.015>
19. International Committee of Medical Journal Editors (2014) Recommendations for the conduct, reporting, editing, and publication of scholarly work in medical journals. Retrieved from www.icmje.org.
20. Kelley, K., & Preacher, K. J. (2012). On effect size. *Psychological Methods*, 17(2), 137–152. <https://doi.org/10.1037/a0028086>
21. Kirk, R. E. (1996). Practical significance: A concept whose time has come. *Educational and Psychological Measurement*, 56(5), 746–759. <https://doi.org/10.1177/0013164496056005002>
22. Matthews, R., Wasserstein, R., & Spiegelhalter, D. (2017). The ASA's p -value statement, one year on. *Significance*, 14, 38–41. <https://doi.org/10.1111/j.1740-9713.2017.01021.x>
23. Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), aac4716. <https://doi.org/10.1126/science.aac4716>
24. Schäfer, T., & Schwarz, M. A. (2019). The Meaningfulness of Effect Sizes in Psychological Research: Differences Between Sub-Disciplines and the Impact of Potential Biases. *Frontiers in psychology*, 10, 813. <https://doi.org/10.3389/fpsyg.2019.00813>
25. Trafimow, D., & Marks, M. (2015). Editorial. *Basic and Applied Social Psychology*, 37(1), 1–2. <https://doi.org/10.1080/01973533.2015.1012991>
26. Valentine, J. C., Aloe, A. M., & Lau, T. S. (2015). Life After NHST: How to Describe Your Data Without “ p -ing” Everywhere. *Basic and Applied Social Psychology* 37(5), 260-273. doi:10.1080/01973533.2015.1060240
27. Wasserstein R. L., & Lazar N. A. (2016). The ASA Statement on p -Values: Context, Process, and Purpose. *The American Statistician*, 70(2), 129–133.

<https://doi.org/10.1080/00031305.2016.1154108>

Wilkinson, L., & Task Force on Statistical Inference, American Psychological Association, Science Directorate. (1999). Statistical methods in

psychology journals: Guidelines and explanations. *American Psychologist*, 54(8), 594–604.

<https://doi.org/10.1037/0003-066X.54.8.594>

Надійшла до редколегії: 18.05.2021

Прийнята до друку: 21.05.2021