

Брайан Резнік, Джулія Белаз

15 НАУКОВИХ ПУБЛІКАЦІЙ ВІДОМОГО ДОСЛІДНИКА КОРНЕЛЬСЬКОГО УНІВЕРСИТЕТУ БУЛО ВІДКЛИКАНО. ЦЕ НЕМАЛО

Брайан Вансінк – повчальна історія про хибні стимули в науці

Переклад здійснив Тимофій Брік за: <https://www.vox.com/science-and-health/2018/9/19/17879102/brian-wansink-cornell-food-brand-lab-retractions-jama>

Одразу шість наукових статей відкликано протягом одного дня – ще й у супроводі з прес-релізом для поширення науковими журналістами всього світу – що це як не страшний сон будь-якого вченого? Саме так вчинила мережа наукових журналів *JAMA* у вересні 2018 р. з Брайаном Вансінком – відомим ученим Корнельського університету. Вансінк був директором Корнельської лабораторії з дослідження їжі та товарних марок¹⁷. Тривалий час Вансінк мав репутацію найкращого спеціаліста щодо патернів поведінки навколо вживання їжі.

Невдовзі після оголошення мережі *JAMA* Корнельський університет заявив, що спеціальна комісія встановила факт порушення наукової доброчесності Вансінком, а також про припинення його контракту з 30 червня 2019 року. Ректор Корнельського університету Майкл Котлікоф заявив, що в проміжку до цього часу Вансінк не буде викладати та займатись дослідженнями. Натомість він буде задійний у процес перегляду попередніх досліджень разом з університетською комісією.

У своєму зверненні до *Vox*¹⁸ сам Вансінк спростував висновки комісії. “Не було жодного шахрайства, жодного навмисного викривлення результатів, плагіату чи привласнення чужих ідей”, – написав він. “Я вірю, що всі мої наукові здобутки будуть або підтвержені, або поглиблені, або модифіковані іншими дослідницькими групами”.

Навіть якщо ви ніколи не чули про Вансінка, ви, скоріш за все, знайомі з його ідеями. Його дослідження, цитовані понад 20 000 разів, описували, як наше середовище формує індивідуальне сприйняття їжі і що ми в кінцевому підсумку споживаємо. Саме його дослідження стало одним із факторів, чому великі виробники почали продавати закуски в невеликих упаковках у 100 калорійних порціях. Колись Вансінк очолював комітет Міністерства сільського господарства США з питань дієтичних рекомендацій та впливав на державні рішення. Він допомагав корпорації Гугл та армії США реалізовувати програми заохочення здорового харчування.

Але за останні кілька років “картковий будинок” його наукової роботи і впливу почав руйнуватися. Група скептично налаштованих дослідників та журналістів, зокрема Стефані Лі BuzzFeed, уважно вивчила доробок групи психологів споживання під керівництвом Вансінка, які працювали в Лабораторії їжі та торгових марок у Корнелі. Виявилось, що недоброчесна робота з даними в цій лабораторії процвітала.

Загалом було відкликано 15 досліджень Вансінка, включаючи шість із мережі JAMA. Серед них були дослідження, які вказували, що голодні люди купують продукти з більшою кількістю калорій; що передзамовлення обіду може допомогти обрати більш здорову їжу і що люди вживатимуть більші порції, якщо серверувати столи тарілками більшого розміру.

У прес-релізі JAMA заявили, що Корнельський університет не зміг “надати гарантій щодо наукової обґрунтованості шести досліджень”, оскільки вони не мали доступу до оригінальних даних Вансінка. Отже, ідеї Вансінка не обов’язково є хибними, скоріш за все він не зміг надати переконливих доказів, щоб їх остаточно підтвердити. Відповідно до заяви Корнельського університету академічна недоброчесність Вансінка включала в себе: “викривлення даних, проблемні статистичні методи, невміння належно документувати та зберігати результати досліджень, а також присвоєння авторства”.

Однак ця історія набагато глибша, ніж просто дії одного окремого дослідника. Ця історія важлива, бо вона проливає світло на систематичну проблему в науці, яка існує в лабораторіях по всьому світу – проблему, що, на думку реформаторів науки, потребує негайного розв’язання. Ось що вам потрібно знати про цю проблему.

П’ятнадцять досліджень Вансінка були відкликані, ще десятки поставлені під сумнів

Вансінк мав хист до досліджень, які були зачіпкою для ЗМІ, включаючи Vox¹⁹. У 2009 році Вансінк і його співавтор опублікували дослідження, яке миттєво стало вірусним. Відповідно до цього дослідження популярні кулінарні книги сприяли збільшенню ваги серед американців²⁰. Дослідники виявили, що рецепти в останніх виданнях кулінарних книг – які продавались накладом понад 18 мільйонів примірників з 1936 року – містили більше калорій і більші розміри порцій порівняно з ранніми виданнями. У ході дослідження було проаналізовано 18 класичних рецептів, які з’явилися у книзі “Задоволення від готування” (Joy of Cooking) з 1936 року. Було виявлено, що з 1936 року середня щільність калорій у цих рецептах зросла на 35 відсотків на порцію.

Інше відоме дослідження Вансінка було присвячене «бездонним мискам». У цьому дослідженні було зроблено висновок, що люди жадібно їстимуть суп, поки вони мають безкоштовні добавки. Його інше дослідження про «поганий попкорн» продемонструвало, що люди можуть переїдати не смачну й неякісну їжу, якщо давати їй величезними порціями.

Усі ці дослідження допомогли Вансінку просунути його цілісну дослідницьку програму – вплив оточення на те, як ми приймаємо рішення про вживання їжі та інші аспекти свого життя.

Утім у 2016 році розпочались критичні запити щодо його публікацій, після того Вансінк опублікував допис у блозі, в якому ненавмисно визнав, що заохочував своїх аспірантів брати участь у сумнівній дослідницькій практиці. Відтоді вчені розпочали прискіпливо вивчати його статті, шукаючи помилки, відхилення й будь-які підозрілі сліди. І десятки дивних випадків були розкриті.

У більш ніж одному випадку Вансінк неправильно вказував вік учасників опублікованих досліджень, змішуючи дітей віком від 8 до 11 років з малюками. Якщо коротко, такі колективні зусилля призвели до створення цілого досьє проблемних висновків у публікаціях Вансінка.

На сьогодні 15 його праць відкликано. І це особливо приголомшливо, враховуючи, що Вансінк був так високо цитований і його робота була настільки впливовою – Вансінк отримував урядові гранти, допомагав формувати маркетингові практики в харчових компаніях та співпрацював з Білим домом, впливаючи на продовольчу політику в цій країні.

Вансінка звинувачували в “очманілій гонитві за значимістю”²¹

Однією з найбільших проблем науки, яку демонструє історія Вансінка, є підхід “публікації – понад усе”.

Щоб більш успішно конкурувати за грантове фінансування, вчені публікуються у шанованих наукових журналах. Щоб їхня робота була прийнята цими журналами, їм потрібні позитивні (тобто статистично значущі) результати.

Це чинить тиск на лабораторії як у випадку Вансінка. Такі лабораторії займаються тим, що відоме як гонитва за значимістю чи “*p*-хакерство”. Латинське *p* вказує на міру статистичної значущості. Як правило, дослідники сподіваються, що їхні результати дають значення *p* менше, ніж 0,05, – межу, за яку вони можуть назвати свої результати статистично значущими.

Значення p дещо складно пояснити. Але в основному вони є інструментом, який допомагає дослідникам зрозуміти, як часто можна отримати такі результати. Якщо результати дослідження можна отримати з невеликою ймовірністю, вчені почуваються впевненіше, що їхня гіпотеза була правильна.

Але диявол у деталях. Значення $p = 0,05$ не так і важко отримати, якщо ви по-різному опрацюєте свої дані й виконуете величезну кількість тестів. Підкидаючи монетку, ви вважаєте, що отримати герб 10 разів поспіль можна дуже рідко. Якщо ви маєте цей результат, то можете подумати, що він не випадковий і монета “мічена”, щоб випадати гербом. Відповідно, ваші результати є статистично значущими.

Але що як ці 10 гербів випали випадково (це може статися), саме в цей момент ви раптом вирішили, що більше монетку підкидати не треба? Якби ви продовжували підкидати її, то припинили б вірити, що монетка “мічена”.

Відповідно, зупиняти експеримент саме в моменти, коли було отримано потрібне значення $p = 0,05$, є прикладом гонитви за значимістю. Однак є й інші способи – наприклад, мати дані щодо великої кількості результатів, але повідомляти лише про ті, що досягають статистичної значущості. Провівши багато статистичних тестів, ви суто випадково знайдете щось значуще.

За словами журналістки Лі з BuzzFeed, яка отримала в своє розпорядження електронні листи Вансінка, замість того, щоб перевіряти гіпотези та повідомляти отримані висновки, Вансінк часто заохочував своїх підлеглих до “викручування” в такий спосіб, щоб отримати цікавіші чи більш бажані результати.

У цілому Вансінк займався підбором потрібних значень p або, як одна із дослідниць зі Стенфорда, Крістін Сайнані сказала BuzzFeed, “очманіла гонитва за значимістю”²².

Недбалість та недоброчесність Вансінка можуть бути більшими, ніж зазвичай у науці. Проте багато інших науковців також визнавали, що вони займались певною формою полювання на значимість.

Опитування 2000 психологів у 2012 році виявило, що тактика гонитви за значимістю є доволі поширеною. П'ятдесят відсотків респондентів визнали, що вони звітували лише про вдалі дослідження, які отримали очікувані результати (ігноруючи інші дані, що не були такими переконли-

вими). Близько 20 відсотків визнали, що припиняли збір даних після того, як отримали результат, на який сподівались. Більшість респондентів вважали, що їхні дії можна виправдати. Багато хто думав, що гонитва за значимістю – це спосіб знайти реальний сигнал у статистичному шумі. Та це неправда. Дедалі частіше навіть усталені істини з підручників виявляються неправдивими після того, як сучасні дослідники їх додатково перевіряють складнішими й надійнішими дослідницькими методами.

Багато людей працюють над тим, щоб зупинити гонитву на значимість

Насправді існує цілий рух учених, які прагнуть побороти недоброчесні практики в науці (наприклад, як ті, в яких звинувачують Вансінка). Разом ці вчені висувають три основні вимоги, які набирають дедалі більшої популярності.

1. Попередня реєстрація дизайну досліджень – це величезна гарантія від гонитви на значимість. Попередня реєстрація означає, що вчені публічно беруться за розроблення експерименту, перш ніж вони почнуть збирати дані. Це значно ускладнює відбір лише потрібних результатів експерименту.
2. Відкритий обмін даними – все частіше вчені закликають своїх колег робити всі дані своїх експериментів доступними для всіх, хто їх може перевірити (звичайно, з винятками для особливо чутливої інформації). Це гарантує, що ті неякісні дослідження, які пройшли рецензування, все одно можуть бути повторно перевірені в майбутньому.
3. Відтворення результатів – учені прагнуть знати: чи можуть бути отримані знову після більш ретельних перевірок ті результати, які вже були опубліковані в наукових журналах. Відповідно, докладається чимало зусиль, щоб із точністю повторити результати попередніх досліджень (точність може бути або методична, або концептуальна).

Є й інші можливі кроки. Окремі вчені систематично закликають до переосмислення терміна статистичної значимості. Інші ж учені стверджують, що будь-яка довільна межа статистичної значимості стимулюватиме вчених до суто формальних процедур.

Тож дедалі частіше вчені звертаються до інших форм математичного аналізу, наприклад, байєсівської статистики, яка дещо інакше поводить себе з даними. Тоді як питання значення p : “Наскільки рідкісні ці числа?”, Байєсівський підхід запитує: “Яка ймовірність, що моя гіпотеза є найкращим поясненням результатів, яких ми досягли?”

Жодне єдине рішення не може бути остаточною панацеєю. Важливо визнати, що наука має розв’язати набагато більш фундаментальну проблему: її культуру.

У 2016 році Vox надіслав опитування більш ніж 200 учених із запитанням: «Якби ви могли змінити одну з засад функціонування науки сьогодні, що це було б і чому?» Одна з постійних тем у цих відповідях була: інститути науки мають навчитись цінувати і невдачі, замість того, щоб схвалювати виключно публікації будь-якою ціною.

Один молодий учений сказав нам: “Я відчуваю, як розриваюсь між питаннями, які, як я знаю, приведуть до статистичної значущості, та питаннями, які мають фундаментальне значення”.

Браян Вансінк зіткнувся з тією ж дилемою. І стає дедалі зрозуміліше, який саме шлях він обрав.