

УДК 001+655

О.І. МРИГЛОД¹, Р. КЕННА², Ю.В. ГОЛОВАЧ¹, Б. БЕРШ³

¹ Інститут фізики конденсованих систем Національної академії наук України
вул. Свенціцького, 1, Львів, 79011, Україна

² Університет Ковентрі, дослідницький центр прикладної математики
Priory St., Coventry, CV1 5FB, United Kingdom

³ Університет Лотарингії
Campus de Nancy, В.Р. 70239, 54506 Vandœuvre lès Nancy Cedex, France

ПРО ВИМІРЮВАННЯ НАУКОВОЇ ЕФЕКТИВНОСТІ

У статті обговорено проблеми оцінювання ефективності наукової діяльності, зокрема роботи наукових колективів. Порівняно два підходи до вирішення цього завдання: експертне оцінювання та використання даних про цитованість опублікованих робіт. З огляду на актуальність дискусії про переваги та недоліки кожного з підходів, здійснено спробу відповісти на запитання: наскільки добре корелюють між собою експертні оцінки та наукометричні індекси й чи можна їх використовувати взаємозамінно?

Ключові слова: наукометрія, ефективність наукових груп.

ВСТУП

Проблема оцінювання наукової праці, либонь, є ровесницею самої науки [1, 2]. Адже від початку зародження впорядкованого критичного мислення робляться спроби виміряти, формалізувати та оцінити одержані результати пізнавальної діяльності. Всі сучасні підходи до оцінювання науки фактично є результатом тривалої еволюції форм і методів самої наукової праці [3]. Усі вони, починаючи від найпростіших методів підрахунку кількості публікацій, числа захищених аспірантів чи виграних грантів і до аналізу ступеня впровадження наукових результатів чи їх експертної оцінки, покликані зрозуміти роль науки і певним чином «вкласти» її у формальні рамки.

Проте жодні з традиційних рамок не придатні для науки, яку можна розглядати як абсолютно іншу, особливу сферу людської діяльності, що поєднує як рутинну працю,

так і елементи творчості. З іншого боку, наука є одним із напрямів державної політики, споживачем бюджетних коштів, тому завдання хоча б часткової її формалізації та оцінювання є дійсно актуальним. Про це свідчить поява все нових рейтингів наукових установ і видань, збільшення кількості опублікованих звітів різних державних і приватних організацій, що займаються вивченням ефективності установ чи країн, а також результатів їх порівняльного аналізу [4, 5].

У світі сучасних інформаційних технологій і засобів автоматичного аналізу великих обсягів бібліографічної інформації за лічені секунди можна підрахувати кількість публікацій чи їх цитувань, провести моментальний статистичний аналіз даних та ще й вивести результати у вигляді наочних діаграм.

Проте спокуса використання автоматичних процедур з метою порівняння продуктивності науковців, дослідницьких груп, установ чи країн приховує ряд небезпек. Навіть скрупульозно сформована бібліографічна база даних не є вичерпним джерелом

інформації, не кажучи вже про цілу низку нюансів, що виникають під час її оброблення. Адже невідомо, яким чином справедливо розподілити умовні бали між співавторами однієї роботи або між кількома установами, до яких одночасно відносить себе один і той самий автор. Так само серед сухої статистики щодо зібраних цитувань важко виділити так звані «негативні цитування», що означають посилення на джерела в контексті їх критики. Неможливо відсіяти частку цитувань, які були просто скопійовані з чужих списків використаних джерел, часто разом із допущеними помилками [6–8]. На сьогодні немає автоматичних процедур, які б дали змогу врахувати ці та інші нюанси. Тому найбільш обґрунтованим і надійним базисом для аналізу наукової діяльності нині залишається експертна оцінка [9, 10].

Незважаючи на численну критику, пов'язану переважно з можливою суб'єктивністю людини-експерта (див., наприклад, [11]), взаємне оцінювання у сфері науки (так зване *peer-review*) традиційно вважають запорукою якості наукових результатів. Одним із найяскравіших прикладів такого роду оцінювання є незалежне рецензування рукописів, що подаються для друку в наукових періодичних виданнях. Той самий принцип використовують під час атестацій та захистів: саме колеги-експерти покликані засвідчити рівень представлених результатів. Однак застосування експертної оцінки для аналізу діяльності великих дослідницьких груп чи установ — справа вельми трудомістка, а отже, затратна як у плані фінансів, так і часу. Саме тому не припиняють звучати заклики до пошуку можливої альтернативи із застосуванням хоча б часткової автоматизації у сфері наукометрії [12, 13].

З огляду на наявну критику як експертного оцінювання, так і автоматичних процедур для одержання наукометричних індикаторів, найбільш виваженим вважають використання комбінованих методів. Сучасні методи автоматичного аналізу даних та кількісні індикатори пропонується залучати як допоміжну ланку в процедурі експертного оцінювання [14].

У цій статті ми порівнюємо результати оцінювання ефективності наукових колективів на основі двох різних підходів: рецензування та аналізу одержаних цитувань. Головне запитання, відповідь на яке шукатимемо в нашому аналізі, — наскільки добре корелюють експертні оцінки та наукометричні індекси та чи можна їх використовувати взаємозамінно? Метою статті є продовження дослідження, розпочатого нами в попередніх роботах [15, 16], а також ознайомлення з ним якнайширшого кола українських науковців. У дослідженні використано як усереднені, або *відносні*, показники ефективності, які можна розглядати як певний коефіцієнт якості групи, так і *абсолютні* значення, пропорційні кількості членів групи. Очевидно, що обидва показники є взаємозалежними: перемноживши відносний показник якості на кількість членів групи, одержимо відповідний абсолютний показник. Обидва значення широко використовують на практиці: пропорційно до абсолютних оцінок якості наукових груп може відбуватися розподіл коштів з боку державних чи інших фондів; відносний же показник якості слугує для порівняння різних наукових колективів та складання рейтингів.

Очевидно, що одержані результати залежатимуть від обраної вибірки даних, адже у різних країнах по-різному організовано наукові дослідження та неоднаково відбувається розвиток окремих галузей науки. Для практичної реалізації наших розрахунків ми обрали дані про роботу наукових груп університетів Великої Британії з огляду на те, що саме у цій країні централізована офіційна процедура аналізу ефективності науки та освіти — Research Assessment Exercise (RAE) [17] — має добре задокументовану традицію, а її результати є публічно доступними. З часу свого заснування система RAE ґрунтується на експертному судженні спеціально обраних груп фахівців. З іншого боку, результати оцінювання тієї самої вибірки наукових груп нам люб'язно надала приватна компанія «Evidence», яка нещодавно ввійшла до складу Thomson Reuters під назвою Research

Analytics [18], а до того самостійно здійснювала на замовлення аналіз діяльності наукових установ.

Отже, надалі ми порівняємо результати аналізу ефективності роботи науковців Великої Британії, проведеного за двома системами оцінювання — RAE та Evidence. Після короткого ознайомлення з відповідними процедурами оцінювання роботи наукових колективів ми наведемо результати аналізу наукової продуктивності для різних галузей науки та розглянемо кореляцію цих оцінок.

ОПИС БРИТАНСЬКОЇ ПРОЦЕДУРИ ОЦІНЮВАННЯ RAE

У 1986 р. у Великій Британії вперше було проведено офіційну процедуру оцінювання ефективності наукових та освітніх закладів — Research Assessment Exercise. Відтоді що три-п'ять років результати такого оцінювання стають основою для рейтингування вищих навчальних закладів та розподілу коштів між ними. Попередня процедура завершилася у 2008 р., а вже нині у британських університетах проводять активну підготовку до наступної, що завершиться у 2014 р. під дещо іншою назвою — Research Excellence Framework та з певними змінами у правилах.

В основу системи RAE покладено незалежну думку кола експертів, відібраних за тематикою. При цьому головний акцент зроблено не на індивідуальному показнику, а на вивченні загальної ефективності цілої наукової групи. Для цього науковий персонал кожної установи, що бажає взяти участь у розподілі державних фондів, групується згідно зі своєю тематикою на так звані модулі — *units of assessment* (UOA). Зокрема, в рамках процедури RAE-2008 було сформовано перелік із 67 тематичних напрямів, які, у свою чергу, розподілено на так звані панелі — *panels*. Для кожної з таких панелей, що позначаються великими латинськими літерами, визначено власне коло експертів і навіть передбачено деякі особливості проведення процедури. Хоча загальна стратегія RAE зберігається для всіх галузей знань, у деяких випадках її можна уточнювати. Адже

навіть така загальноприйнята форма вираження наукового результату, як статті, опубліковані у науковій періодиці, для деяких наукових дисциплін є менш значущою, ніж, скажімо, монографії (історичні науки), підручники (педагогіка) чи музичні твори (музика). Отже, групування окремих дисциплін у тематичні панелі надає всій процедурі більшої гнучкості без втрати універсальності. Наведемо приклади тематичних панелей.

Панель F:

- загальна математика (Pure Mathematics, UOA 20);
- прикладна математика (Applied Mathematics, UOA 21);
- статистика та дослідження операцій (Statistics and Operational Research, UOA 22);
- комп'ютерні науки та інформатика (Computer Science and Informatics, UOA 23).

Панель H:

- архітектура та антропогенне середовище (Architecture and the Built Environment, UOA 30);
- міське та сільське планування (Town and Country Planning, UOA 31);
- географія та навколишнє середовище (Geography and Environmental Studies, UOA 32);
- археологія (Archaeology, UOA 33).

Для того щоб отримати якісну характеристику від експертів RAE, кожна установа має подати заявку згідно з оприлюдненими вимогами. Як правило, кожне таке подання містить велику кількість заповнених форм, у тому числі дані про науковців, що входять до тематичних модулів. Згідно з правилами, кожний дослідник має представити 4 власні публікації за відповідний період. Якщо публікація не є одноосібною, то додатково зазначають внесок автора у спільну роботу. За потреби можна вказати й іншу додаткову інформацію, важливу для подальшої процедури. Експерти відповідної тематичної панелі вивчають усі подані заявки і генерують так званий якісний профіль для певної наукової групи, що визначає, яка частка проаналізованих результатів групи припадає на кожний із 5 рівнів якості:

- 4* — найвищий рівень якості у світовому масштабі;

- 3* – високий міжнародний рівень якості;
- 2* – міжнародний рівень;
- 1* – національний рівень;
- Unclassified – низький рівень якості або невідповідність опублікованим критеріям RAЕ.

Підсумковий якісний профіль групи містить три підпрофілі: *Output* (75%), *Environment* (20%) та *Esteem* (5%). Пропорції підпрофілів можуть дещо варіюватися для різних панелей. Підпрофіль *Output* характеризує якість опублікованих результатів, ґрунтуючись на оцінюванні поданих у заявці 4 статей на кожного члена групи. Для формування підпрофілю *Environment* експерти оцінюють інфраструктуру кожної установи, роботу зі студентами, політику щодо персоналу, загальну стратегію тощо. Підпрофіль *Esteem* дає змогу врахувати всі нагороди, відзнаки, участь у комітетах чи редакційних колегіях та інші особисті заслуги кожного із членів групи.

Приклад підсумкового профілю якості та відповідних підпрофілів для одного з університетів Великої Британії наведено у табл. 1. Алгоритм формування загального профілю, описаний у документації RAЕ на офіційному веб-сайті [17], містить процедуру округлення та врахування кожного з підпрофілів відповідно до заданої пропорції.

Для того щоб мати змогу коректно порівняти результати RAЕ з наукометричними показниками на основі цитованості, замість загального профілю якості ми використовуємо лише підпрофіль *Output*. Так само, як і аналіз цитувань, лише він ґрунтується на оцінюванні власне рівня публікацій.

Крім очевидних переваг, які дає профіль якості RAЕ в аспекті престижу чи маркетингу, він визначає розмір фінансування, яке отримуватиме від держави кожна з груп установи впродовж кількох років. Розподіл коштів щороку визначається за спеціальною формулою, яку оголошує Спеціальна рада Англії з питань фінансування вищої освіти* – Higher

* У Великій Британії немає чіткого поділу на навчальні та дослідницькі заклади – за рідкісними винятками, усі наукові дослідження виконують у вищих навчальних закладах.

Табл. 1.

Профіль якості та його підпрофілі для групи біології Університету Йорка [17]

Тип профілю	Відсоток дослідницької активності кожного рівня якості				
	4*	3*	2*	1*	Unclassified
Output	14,7	37,5	36,5	10,4	0,9
Environment	55,9	41,0	3,1	0,0	0,0
Esteem	8,2	24,6	36,2	27,1	3,9
Загальний	25,0	35,0	30,0	10,0	0,0

Education Funding Council for England (HEFCE). Зокрема, безпосередньо після завершення процедури RAЕ-2008 використовували таку формулу розподілу фондів:

$$s = p_{4*} + \frac{3}{7} p_{3*} + \frac{1}{7} p_{2**}, \quad (1)$$

де p_{n*} – значення із загального профілю якості, що припадає на рівень n , а s можна розглядати як відносну міру якості наукових досліджень у певній установі та в рамках певної тематики. Саме цю формулу застосовуємо далі для розрахунку відносного показника якості, лише використовуючи значення із підпрофілю *Outputs* p_{n*} замість значень із загального профілю якості. За термінологією попередніх робіт [19, 20], цей відносний показник якості s надалі називатимемо просто *якістю* групи. Відповідно, абсолютний показник якості позначимо великою літерою S і називатимемо *силою* групи:

$$S = sN, \quad (2)$$

де N – кількість членів групи.

Нині RAЕ та подібні процедури оцінювання ефективності наукової діяльності в інших країнах, що ґрунтуються на незалежному експертному оцінюванні, вважають найнадійнішими. Проте, очевидно, що їм властиві всі недоліки рецензування, згадані вище. Крім того, масштабна підготовка та проведення такого роду заходів, без сумніву, впливають на нормальну роботу наукових установ, тим самим спотворюючи результат. Такий феномен відомий під назвою «закон

Гудхарта» (Goodhart's law), що у спрощеному вигляді формулюється так: коли якась міра сама по собі стає метою, вона перестає бути надійною мірою. Так само, як існують певні організаційні заходи, що дають змогу дещо підвищити імпаکت-фактор видання, так фактично кожен вищий навчальний заклад Великої Британії намагається заздалегідь забезпечити собі високий результат RAE: певною мірою це стає однією із цілей. Інший, вельми вагомий, недолік експертної масштабної процедури — її вартість. З одного боку, затратною у фінансовому плані є сама робота експертів. З іншого боку, цих експертів залучають саме з наукової сфери, тобто значну частку свого робочого часу вони присвячують додатковим обов'язкам.

З огляду на викладене вище, зрозуміло, що періодично з'являються дослідження, які порівнюють рейтинги наукових груп за версією RAE і рейтинги, що ґрунтуються на різноманітних наукометричних індикаторах на основі даних про цитування.

THOMSON REUTERS RESEARCH ANALYTICS

Як уже згадувалося, для того щоб порівняти результати RAE для наукових колективів із показниками їх цитованості, зручно було скористатися даними приватної компанії «Evidence» (нині Thomson Reuters Research Analytics). Для порівняння використовували значення так званого показника нормалізованого впливу цитувань, *normalised citation impact* (NCI), для кожної наукової групи. Значення NCI (далі позначаємо його малою літерою i) є відносним. Відповідно, можна розрахувати абсолютне значення ефективності:

$$I = iN, \quad (3)$$

де N — кількість членів групи. Оскільки наукові цитування лише опосередковано свідчать про якість наукових результатів, а більшою мірою — про їх вплив на наступні дослідження, то, оперуючи індикаторами на основі цитованості, вживатимемо терміни *середній вплив* та *загальний вплив* (i та I) за аналогією до *якості* та *сили* (s та S).

Для свого аналізу компанія «Evidence» використовувала дані про цитування з бази Thomson Reuters. Головною перевагою індикатора NCI є нетривіальна процедура нормування, що дає змогу компенсувати відмінності в цитованості різних галузей науки, різних видань чи навіть різних проміжків часу. Загальну кількість цитувань кожної статті ділять на середню кількість цитувань для певного видання і в певний рік. У процесі розрахунку NCI для колективу науковців нормовані значення цитованості для кожної статті, внесеної до заявки в RAE-2008, сумують та усереднюють за кількістю членів групи.

ПОРІВНЯННЯ СИСТЕМ ОЦІНЮВАННЯ

Наша мета — порівняти між собою відносні показники s та i , що характеризують *якість* і *середній вплив* наукових груп, а також абсолютні показники S та I , що свідчать про *силу* та *загальний вплив* колективів загалом. Як показано далі, на рівні абсолютних показників кореляція є достатньо сильною, тоді як значення відносних показників, а також рейтинги установ, що ґрунтуються на цих відносних показниках, істотно різняться. Ці висновки переважно зберігаються для різних дисциплін та не залежать від розмірів групи. При цьому під розміром групи маємо на увазі її належність до категорії великої, середньої чи малої. Проблему значення розміру для наукового колективу детально було досліджено в роботах [19, 20], де вперше поставлено питання про значення розміру наукового колективу для його ефективності. Виявилось, є певна залежність *якості* групи s від кількості її членів N : доки науковий колектив збільшується до певного характерного розміру, доти відносний показник його ефективності зростає; коли ж група перевищує певне значення N , поява нових членів не підвищує *якості* групи. Відповідно, залежність *сили* групи S від N не є тривіально лінійною: насправді можна побачити, що збільшення розміру групи дає найбільший приріст її *сили* лише за певних значень $N_k < N < N_c$. Отже, було визначено три категорії наукових груп: малі ($N \leq N_k$),

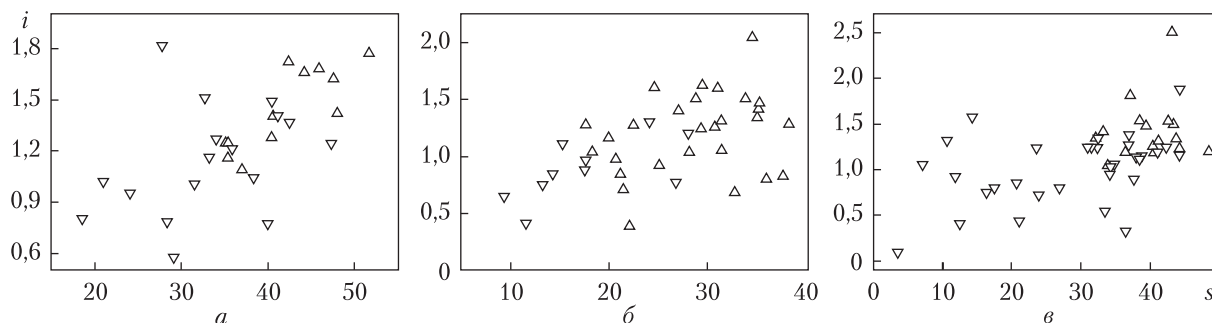


Рис. 1. Залежність між відносними показниками s та i для наукових груп у галузі хімії (а), соціології (б) та географії (в). Трикутниками вершиною донизу показано результати для малих і середніх груп, вершиною догори — для великих груп [15]

середні ($N_k \leq N \leq N_c$) та великі ($N \geq N_c$). Порогові значення змінюються залежно від галузі науки, скажімо, для хімії $N_k \approx 18$ та $N_c \approx 36$, для соціології $N_k \approx 7$ та $N_c \approx 14$, а для географії $N_k \approx 15$ та $N_c \approx 30$. Було показано, що найвищий приріст ефективності дає збільшення кількості членів середньої наукової групи [19, 20]. Зважаючи на можливі відмінності для груп різних розмірів, у нашій роботі окремо виділено великі групи та сукупність малих і середніх (розглядаємо їх разом, оскільки, як правило, великих груп є набагато більше, ніж малих і середніх). Слід зауважити, що для різних галузей науки співвідношення груп різного розміру не є однаковим, хоча майже в усіх випадках кількість саме великих колективів найбільша.

Якби показник *середнього впливу* наукової групи, розрахований на основі аналізу цитувань, виявлявся тотожним значенню *якості*, що визначається експертами, можна було б сміливо заявляти про можливість заміщення трудомісткої процедури рецензування відповідним індексом на основі цитувань. У такому разі кореляція між значеннями s та i була б строго лінійною. Для перевірки цієї гіпотези ми обрали три дисципліни: хімію та соціологію, які представляють точні й суспільні науки, а також географію, що в інтерпретації RAE поєднує в собі природничий і суспільний аспекти. Таким чином, аналізувалися дані про наукові групи, що ввійшли до таких модулів:

- Хімія (Chemistry, UOA 18);
- Соціологія (Sociology, UOA 41);
- Географія та вивчення навколишнього середовища (Geography and Environmental Studies, UOA 32).

Відповідні графіки кореляцій s та i для вибраних дисциплін із врахуванням розміру наукових груп зображено на рис. 1. Очевидно, що кореляція в усіх трьох випадках є далекою від лінійної. Зберігається лише тенденція до позитивної кореляції, коли більше s відповідає також більшому i , проте самі значення досить розкидані. Найпростішим методом перевірки якості лінійної апроксимації є розрахунок коефіцієнта Пірсона: чим ближче його абсолютне значення до 1, тим краще точки «лягають» на пряму. Розраховані значення коефіцієнта Пірсона, наведені у табл. 2, свідчать про переважно слабку кореляцію між досліджуваними величинами. Цікаво, що для деяких дисциплін відносні показники ефективності груп корелюють сильніше для груп великого розміру (хімія), тоді як для інших дисциплін кращі показники кореляції спостерігаються для малих і середніх груп.

Побудувавши рейтинг установ на основі s , ми бачимо, що він значно відрізняється від аналогічного рейтингу на базі i . Використаємо спеціальний коефіцієнт Спірмена, що дає змогу порівняти рейтинги між собою. Так само, як і для коефіцієнта Пірсона, абсолютне значення коефіцієнта Спірмена, близьке до 1, означає максимальну подібність між двома рейтингами. Розраховані

значення свідчать про те, що рейтинги на основі s та i подібні лише частково — дещо краще для хімії і гірше для соціології та географії (табл. 2). Такі результати суперечать деяким дослідженням, що демонструють високий рівень кореляції між рейтингами на основі даних RAE та на основі підрахунку цитувань. Наприклад, було знайдено від-

Табл. 2.
Значення коефіцієнтів кореляції між відносними показниками s та i для наукових колективів Великої Британії з різних дисциплін

Дисципліна	Коефіцієнт Пірсона			Коефіцієнт Спірмена
	усі групи	великі групи	малі й середні групи	
Хімія	0,60	0,82	0,34	0,62
Соціологія	0,49	0,29	0,64	0,47
Географія	0,51	0,13	0,42	0,47

Табл. 3.
Значення коефіцієнтів кореляції між абсолютними показниками S та I наукових колективів Великої Британії для різних дисциплін

Дисципліна	Коефіцієнт Пірсона		
	усі групи	великі групи	малі й середні групи
Хімія	0,96	0,96	0,79
Соціологія	0,88	0,82	0,73
Географія	0,92	0,56	0,93

Табл. 4.
Значення коефіцієнтів кореляції між абсолютними та відносними показниками ефективності наукових колективів Великої Британії для різних дисциплін

Дисципліна	Коефіцієнт Пірсона	
	для відносних показників s та i	для абсолютних показників S та I
Біологія	0,60	0,97
Фізика	0,48	0,96
Інженерія	0,34	0,92
Історія	0,34	0,88

носно високі значення коефіцієнта Спірмена для деяких специфічних (з погляду методології оцінювання) дисциплін: 0,80 для музики [21] та 0,81 для археології [22].

У нашому дослідженні можна було очікувати доброї кореляції між рейтингами з огляду на використання більш тонкої методики: урахування окремих підпрофілів, що стало можливим лише для результатів RAE-2008, та продуманий алгоритм нормування для розрахунку значень NCI, що є перевагою порівняно з простим підрахунком цитувань. Проте виявилось, що навіть не тривіальна процедура врахування даних про цитування не дає змоги відтворити результат рейтингування наукових груп на основі експертних оцінок. Тому нормалізований показник впливу на основі цитувань не можна назвати альтернативою рецензуванню.

Тепер порівняємо абсолютні показники: *загальний вплив* наукової групи I та її *силу* S . Знову ж таки, якби результат рецензування повністю збігався з наукометричними індикаторами на основі цитувань, ці дві величини були б лінійно взаємозалежними. На рис. 2 показано, як насправді співвідносяться S та I для трьох обраних дисциплін. Бачимо, що перемноживши відносні значення i та s на розмір групи N (формули (2) і (3)), ми одержали досить добре скорельовані абсолютні показники. В усіх трьох випадках чітко видно, що значення дещо розкидані, проте зосереджені в околі прямої лінії. Це спричинено тим, що показники ефективності груп ніби набувають ваги, пропорційної кількості членів групи: показники великих груп (як статистично надійніші) тепер мають більше значення, ніж показники малих груп. У табл. 3 наведено відповідні значення коефіцієнта лінійної кореляції Пірсона, що підтверджують значну лінійну скорельованість величин S та I . З огляду на те, що абсолютні показники не резонно використовувати для складання рейтингів, немає сенсу обраховувати коефіцієнти Спірмена.

Нормовані, або ж відносні, показники ефективності наукових груп дають змогу порівняти середню результативність груп

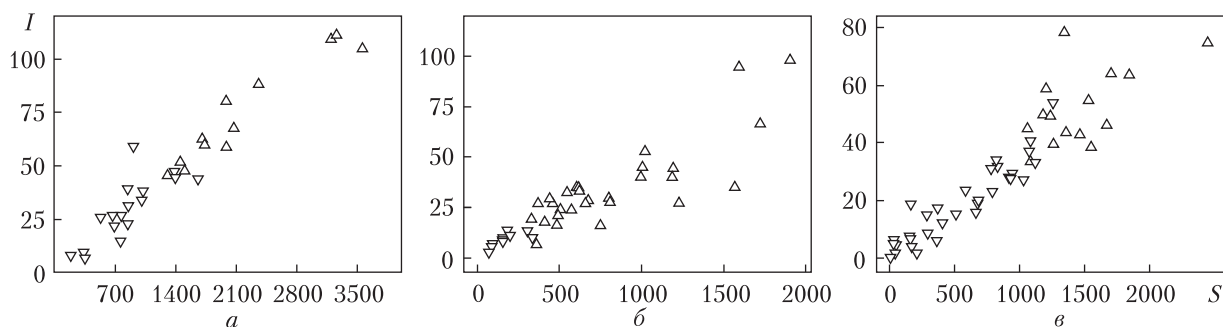


Рис. 2. Залежність між абсолютними показниками I та S для наукових груп у галузі хімії (a), соціології (b) та географії (v). Трикутниками вершиною донизу показано результати для малих і середніх груп, трикутниками вершиною догори — для великих груп [15]

різного розміру. Скажімо, за результатами RAE-2008 приблизно однаковий рейтинг отримали групи з історії з двох різних установ — Відкритого університету (Open University) та Університету Гламоргану (University of Glamorgan) при тому, що розміри цих наукових колективів відрізняються більше, ніж утричі. Проте деякі завдання, такі як розподіл ресурсів, виконуються пропорційно до *сили* групи і тому враховують *абсолютні* показники ефективності. Результати порівняння свідчать про те, що *сила* групи більш-менш корелює з її *загальним впливом*. Тому в деяких випадках для економії часу та зусиль замість процедури експертного оцінювання можна було б використати наукометричні показники на основі аналізу цитувань. Для хімії та соціології такий висновок навіть більш справедливий щодо великих груп (табл. 3), і це зрозуміло на інтуїтивному рівні: більша статистика — менші відхилення і надійніший результат. Про причини кращої кореляції для малих і середніх наукових груп у галузі географії можна лише припускати. Зокрема, важливою відмінністю географії в контексті нашого дослідження є значний міждисциплінарний аспект. За визначенням RAE [17], до географічного модуля ввійшли як дослідження природних явищ навколишнього середовища, так і людської діяльності в цьому контексті.

Для перевірки висновку про суттєву кореляцію між показниками *сили* S та *загального впливу* I наукових груп ми розрахували також ко-

ефіцієнти лінійної кореляції для інших дисциплін: біології, фізики, інженерії та історії. Результати, наведені у табл. 4, свідчать про слабку кореляцію між відносними показниками s та i для всіх дисциплін. Висновок про високу кореляцію між абсолютними значеннями S та I також підтверджується в усіх випадках. Зауважимо, що найкраща кореляція (з коефіцієнтом Пірсона понад 0,90) спостерігається для дисциплін, ближчих до точних наук (hard sciences, за однією з класифікацій в англійській літературі) — біології, фізики та хімії. Так само сильною, хоча з меншим коефіцієнтом кореляції (меншим за 0,90), є кореляція між S та I для представників гуманітарної сфери (soft sciences) — історії та соціології. Відповідні значення коефіцієнтів кореляції для міждисциплінарної географії та прикладної інженерії знаходяться десь посередині, в околі 0,92.

КОРОТКІ ВИСНОВКИ

Підсумовуючи, можна констатувати, що на сьогодні немає стовідсоткової альтернативи процедурі експертного оцінювання у сфері науки. Незважаючи на всі недоліки, незалежне рецензування залишається найкращим гарантом якості наукової роботи. Проте не виключена можливість використання сучасних досягнень наукометричної науки та інформаційних технологій для підтримки прийняття рішень людиною-експертом.

Різного роду помилки, недоліки процедури автоматичного опрацювання даних, різноманітні нюанси, пов'язані з цитуванням, —

усе це зумовлює високий рівень похибки при намаганні оцінити наукову працю індивіда чи цілого колективу. Тому цифри, одержані в такий спосіб, слабко корелюють з оцінками незалежних експертів. Однак якщо оцінювати загальну ефективність усієї наукової групи, то статистична похибка зменшується тим відчутніше, чим більшим є розмір групи. Тому такі зважені оцінки на основі тих самих наукометричних індикаторів дають змогу з високою ймовірністю передбачити експертну оцінку загальної сили групи. Ці висновки справедливі для різних галузей знань, хоча найкраще виконуються для точних наук. Отже, можна сформулювати дві головні тези:

Оцінки рецензентів та індикатори на основі аналізу цитувань, використані для оцінювання середньої ефективності наукових груп, корелюють слабо. Тому не можна застосовувати наукометричні індикатори для формування рейтингів наукових колективів чи установ без урахування думки експертів.

Добре корелюють між собою оцінки рецензентів та індикатори на основі аналізу цитувань, використані для оцінювання загальної ефективності наукових груп. Отже, наукометричні індикатори з певною часткою похибки можна використовувати для виявлення сили або абсолютної ефективності наукових груп.

Дослідження проведено в межах проектів: «Динаміка і кооперативні явища в складних фізичних і біологічних середовищах» (7 Рамкова угода, FP7-PEOPLE, IRSES project N269139) та «Статистична фізика у різноманітних реалізаціях» (7 Рамкова угода, FP7-PEOPLE, IRSES project N295302).

СПИСОК ЛІТЕРАТУРИ

1. Основы науковедения / ред. Н. Стефанов, Н. Яхилел, Я. Фаркаш и др. — М. : Наука, 1985. — 431 с.
2. Жмудь Л.Я. Зарождение истории науки в Античности. — СПб: Изд. Рус. христ. гум. ин-та, 2002. — ISBN 5-88812-172-X.
3. Добров Г.М. Наука о науке. — К. : Наук. думка, 1989. — ISBN 5-12-001112-8.
4. Butler D. University rankings smarten up // Nature. — 2010. — V. 464. — P. 16–17.
5. Williams R., de Rassenfosse G., Jensen P., Marginson S. U21 Ranking of National Higher Education Systems. — University of Melbourne, 2012.
6. Bibliometric evaluation and international benchmarking of the UKs physics research. — Thomson Reuters, 2012.
7. Simkin M.V., Roychowdhury V.P. Stochastic modeling of citation slips // Scientometrics. — 2005. — V. 62, N 3. — P. 367–384.
8. Simkin M.V., Roychowdhury V.P. Read before you cite! // Ithaca: Cornell University Library, 2002.
9. van Raan A.F.J. Fatal attraction: Conceptual and methodological problems in the ranking of universities by bibliometric methods // Scientometrics. — 2005. — V. 62, N 1. — P. 133–143.
10. Derrick G.E., Haynes A., Chapman S., Hall W.D. The Association between Four Citation Metrics and Peer Rankings of Research Influence of Australian Researchers in Six Fields of Public Health // PLoS ONE. — 2001. — V. 6. — e18521.
11. Borrmann L. The Hawthorne effect in journal peer review // Scientometrics. — 2012. — V. 91. — P. 857–862.
12. Science Metrics // Nature. — <http://www.nature.com/metrics>.
13. Warner J. Citation Analysis and Research Assessment in the United Kingdom // B. Am. Soc. Inform. Sci. Tech. — 2003. — V. 30, N 1. — P. 26–27.
14. De Bellis N. Bibliometrics and citation analysis: from the science citation index to cybermetrics. USA. — Scarecrow Press, 2009.
15. Mryglod O., Kenna R., Holovatch Yu., Berche B. Absolute and specific measures of research group excellence // Scientometrics. — 2013. — V. 95, N 1. — P. 115–127.
16. Mryglod O., Kenna R., Holovatch Yu., Berche B. Comparison of a citation-based indicator and peer review for absolute and specific measures of research-group excellence // Scientometrics. — arXiv:1305.6256 [cs.DL].
17. RAE 2008. — <http://www.rae.ac.uk/>.
18. Thomson Reuters Research Analytics. — <http://www.evidence.co.uk>.
19. Kenna R., Berche B. Critical mass and the dependency of research quality on group size // Scientometrics. — 2010. — V. 86, N 2. — P. 527–540.
20. Kenna R., Berche B. Critical masses for academic research groups and consequences for higher education research policy and management // Higher Education Management and Policy. — 2011. — V. 23, N 3. — P. 1–21.
21. Oppenheim C., Summers M.A.C. Citation counts and the Research Assessment Exercise, part VI. Unit of assessment 67 (music) // Information Research. — 2008. — V. 13, N 2. — P. 342.
22. Norris M., Oppenheim Ch. Citation counts and the Research Assessment Exercise. V Archaeology and the 2001 RAE // Journal of Documentation. — 2003. — V. 59, N 6. — P. 709–730.

Стаття надійшла 20.05.2012 р.

О. Мрыглад¹, Р. Кенна², Ю. Головач¹, Б. Бэрш³

¹ Інститут фізики конденсованих систем
НАН України

ул. Свенцицкого, 1, Львов, 79011, Украина

² Университет Ковентри, исследовательский центр
прикладной математики

Priory St., Coventry, CV1 5FB, United Kingdom

³ Университет Лотарингии

Campus de Nancy, B.P. 70239, 54506 Vandœuvre lès
Nancy Cedex, France

ОБ ИЗМЕРЕНИИ НАУЧНОЙ ЭФФЕКТИВНОСТИ

В статье обсуждаются проблемы оценивания эффективности научной деятельности, в частности работы исследовательских коллективов. Сравниваются два подхода к решению этой задачи: экспертные оценки и использование данных о цитировании опубликованных работ. В связи с актуальностью дискуссии о преимуществах и недостатках каждого из подходов делается попытка дать ответ на вопрос: насколько хорошо коррелируют между собой экспертные оценки и наукометрические индексы и взаимозаменяемы ли они.

O. Mryglod¹, R. Kenna², Yu. Holovatch¹, B. Berche³

¹ Institute for Condensed Matter Physics
of the National Academy of Sciences of Ukraine
1 Svientsitskii St., 79011, Lviv, Ukraine

² Coventry University,

Applied Mathematics Research Centre

Priory St., Coventry, CV1 5FB, United Kingdom

³ Université de Lorraine

Campus de Nancy, B.P. 70239, 54506 Vandœuvre lès
Nancy Cedex, France

ON THE PROBLEM OF SCIENCE EVALUATION

We discuss problems of scientific evaluation, making particular emphasis on research group evaluation. In order to perform such an assessments two basic approaches could be used: expert judgments and analysis of the citation history of publications. Each of these approaches has its advantages and drawbacks. Here we make a comparison and address the question of whether a set of automated, scientometric or bibliometric indicators is a suitable substitute for, or component of, peer-review at the level of the research group or department.