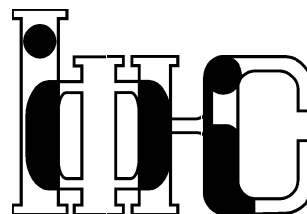


Препринти Інституту фізики конденсованих систем НАН України розповсюджуються серед наукових та інформаційних установ. Вони також доступні по електронній комп'ютерній мережі на WWW-сервері інституту за адресою <http://www.icmp.lviv.ua/>

The preprints of the Institute for Condensed Matter Physics of the National Academy of Sciences of Ukraine are distributed to scientific and informational institutions. They also are available by computer network from Institute's WWW server (<http://www.icmp.lviv.ua/>)

Національна академія наук України



ІНСТИТУТ  
ФІЗИКИ  
КОНДЕНСОВАНИХ  
СИСТЕМ

Олеся Ігорівна Мриглод  
Ральф Кенна  
Юрій Васильович Головач  
Бертран Берш

ПРО ВИМІРЮВАННЯ НАУКОВОЇ ЕФЕКТИВНОСТІ

Роботу отримано 21 червня 2013 р.

Затверджено до друку Вченою радою ІФКС НАН України

Рекомендовано до друку відділом статистичної теорії  
конденсованих систем

Виготовлено при ІФКС НАН України

© Усі права застережені

ICMP-13-02U

О.Мриглод, Р.Кенна, Ю.Головач, Б.Берш

ПРО ВИМІРЮВАННЯ НАУКОВОЇ ЕФЕКТИВНОСТІ

ЛЬВІВ

УДК: 001.893, 316.74:001, 519.24, 53.08

PACS: 01.30.-y, 02.50.-r

### Про вимірювання наукової ефективності

О.Мриглод, Р.Кенна, Ю.Головач, Б.Берш

**Анотація.** У статті обговорюються проблеми оцінювання ефективності наукової діяльності, а зокрема – роботи наукових колективів. Порівнюються два підходи до вирішення цього завдання: експертного оцінювання та використання даних про цитованість опублікованих робіт. З огляду на актуальність дискусії про переваги та недоліки кожного із підходів, в даній роботі здійснено спробу відповісти на запитання: “наскільки добре корелюють експертні оцінки та наукометричні індекси, та чи можна їх використовувати взаємозамінно?”.

### On the problem of science evaluation

O.Mryglod, R.Kenna, Yu.Holovatch, B.Berche

**Abstract.** 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.

Подається в Вісник НАН України

Submitted to Bull. Nat. Acad. Sci. of Ukraine

### Вступ

Проблема оцінювання наукової праці, либонь, є ровесницею самої науки [1, 2]. Адже від початків 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 ґрантів і до аналізу ступеня впровадження наукових результатів чи їх експертної оцінки, – покликані зрозуміти роль науки та певним чином вкласти її у формальні рамки. Проте жодні із традиційних рамок не підходять 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 напрямків державної політики, споживачем бюджетних коштів, тому задача хоча б часткової її формалізації та оцінювання є дійсно актуальною. Про це свідчить поява все нових рейтингів наукових установ та видань, збільшення кількості опублікованих звітів різних приватних та державних організацій, що займаються вивченням ефективності установ чи країн, а також – результатів їх порівняльного аналізу [4, 5].

У світі сучасних інформаційних технологій та засобів для автоматичної обробки великих обсягів бібліографічної інформації, за кілька секунд можна підрахувати кількість публікацій чи одержаних ними цитувань, провести миттєву статистичну обробку даних та ще й вивести у формі привабливих діаграм. Проте спокуса використання автоматичних процедур із метою порівняння продуктивності науковців, дослідницьких груп, установ чи країн приховує ряд небезпек. Навіть скурпульозно сформована бібліографічна база даних не є вичерпним джерелом інформації, не кажучи вже про цілу низку нюансів, що виникають при її обробці. Адже невідомо, яким чином справедливо розподілити умовні “бали” між співавторами однієї роботи, або між кількома установами, до яких одночасно відносить себе один і той же автор. Так само серед сухої статистики щодо зібраних цитувань важко виділити так звані “негативні цитування”, що означають поклики на джерела в контексті їх критики. Неможливо відсіяти частку цитувань, які були просто скопійовані із чужих списків використаних джерел, часто разом із допущеними помилками [7, 8]. На даний момент не існує автоматичних процедур, які б дали змогу врахувати ці та інші нюанси. Тому найбільш обґрунтова-

ним та надійним базисом для аналізу наукової діяльності нині залишається експертна оцінка [9, 10]. Незважаючи на численну критику, в основному пов'язану із можливою суб'єктивністю людини-експерта (див., наприклад, [11]), взаємне оцінювання у сфері науки (так зване реєг-ревію, з англ.) традиційно вважається запорукою якості наукових результатів. Одним із найяскравіших прикладів такого роду оцінювання є незалежне рецензування рукописів, що подаються для друку у наукових періодичних виданнях. Той самий принцип використовується при атестаціях та захистах наукових знань: саме колеги-експерти покликані засвідчити рівень представлених результатів. Проте застосування експертної оцінки для аналізу діяльності великих дослідницьких груп чи установ – справа вельми трудоемка, а отже затратна як у плані фінансів, так і часу. Саме тому не припиняють звучати заклики про пошук можливої альтернативи із застосуванням хоча б часткової автоматизації у сфері наукометрії [12, 13].

З огляду на існуючу критику як експертного оцінювання, так і автоматичних процедур для одержання наукометричних індикаторів, найбільш виваженим вважається використання комбінованих методів. Сучасні методи автоматичної обробки даних та кількісні індикатори пропонується залучати в якості допоміжної ланки у процедурі експертного оцінювання [14].

Порівняємо результати оцінювання ефективності наукових колективів на основі двох різних підходів: рецензування та аналізу одержаних цитувань. Головним запитанням, відповідь на яке шукатимемо в нашому аналізі є: “Наскільки добре корелюють експертні оцінки та наукометричні індекси, та чи можна їх використовувати взаємозамінно?”. Метою цієї статті є продовжити дослідження, розпочаті нами в недавніх роботах [15, 16] а також ознайомити з ними ширше коло українських науковців. У дослідженні використано як усереднені або *відносні* показники ефективності, що можуть розглядатися як певний коефіцієнт якості групи, так і *абсолютні* значення, пропорційні до кількості членів групи. Очевидно, що обидва показники, абсолютний та відносний, є взаємозалежними: перемноживши відносний показник якості на кількість членів групи, одержимо відповідний абсолютний показник. Обидва значення широко використовуються на практиці: пропорційно до абсолютних оцінок якості наукових груп може відбуватися розподіл коштів з боку державних чи інших фондів; відносний же показник якості вживають при порівнянні різних наукових колективів та побудові рейтингів.

Очевидно, що одержані результати будуть залежати від обраної вибірки даних, адже у різних країнах по-різному організовано нау-

кові 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, результати якої є публічно доступними. З часу свого започаткування система RAE базується на експертному судженні спеціально обраних груп спеціалістів. 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 аналіз діяльності наукових установ на замовлення. Отже, надалі у цій статті ми порівняємо результати аналізу ефективності роботи науковців Великої Британії проведені за двома системами оцінювання: RAE та “Evidence”. Після короткого ознайомлення з процедурами оцінювання роботи наукових колективів ми приведемо результати аналізу наукової продуктивності для різних галузей науки та розглянемо кореляцію цих оцінок.

## 1. Опис британської процедури оцінювання RAE

У 1986 році 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, а вже нині в британських університетах проводиться активна підготовка до наступної, що відбудеться у 2014 році під дещо зміненою назвою – Research Excellence Framework – та з певними змінами у правилах.

В основі системи RAE лежить незалежна думка кола експертів, відібраних за тематикою. При цьому головний акцент робиться не на індивідуальний показник, а на вивчення загальної ефективності цілої наукової групи. Для цього науковий персонал кожної установи, що бажає взяти участь у розподілі державних фондів, групується згідно своєї тематики на так звані модулі – “units of assessment” (UOA). Зокрема, в рамках останньої, завершеної у 2008 році процедури RAE було сформовано перелік із 67 тематичних напрямків, які, в свою чергу, погруповані у так звані панелі – “panels”. Для кожної із таких панелей, що позначаються великим латинськими літе-

рами, визначене власне коло експертів і навіть передбачено деякі особливості проведення процедури. Хоча загальна стратегія RAE зберігається для усіх галузей знання, у деяких випадках її можна уточнювати. Адже навіть загальноприйнята форма вираження наукового результату – опубліковані статті у науковій періодиці – для деяких наукових дисциплін є менш значущими, ніж, скажімо монографії (історичні науки), підручники (педагогіка) чи музичні твори (музика). Таким чином, групування окремих дисциплін у тематичні панелі робить усю процедуру більш гнучкою без втрати універсальності. Приклади тематичних панелей:

#### Панель “F”:

- Загальна математика (Pure Mathematics, UOA 20);
- Прикладна математика (Applied Mathematics, OUA 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 одержати якісну характеристику від експертів RAE, кожна установа повинна подати заявку згідно з оприлюдненими вимогами. Як правило, кожне таке подання містить велику кількість заповнених форм, включаючи дані про науковців, що входять до тематичних модулів. Згідно із правилами, кожен дослідник повинен представити свої 4 публікації за відповідний період. Якщо публікація не є одноосібною, то додатково зазначається внесок даного автора у загальну роботу. При потребі, може вказуватися й інша додаткова інформація, що може бути важливою для подальшої процедури. Експерти відповідної тематичної панелі вивчають всі подані

заявки та генерують так званий якісний профіль для кожної наукової групи. Він визначає, яка частка проаналізованих результатів групи припадає на кожен 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 RAE.

Результуючий якісний профіль групи включає у себе три підпрофілі: “Output” (75%), “Environment” (20%) та “Esteem” (5%). Пропорції включення підпрофілів можуть злегка варіюватися для різних панелей. Підпрофіль “Output” характеризує якість опублікованих результатів, базуючись на оцінюванні поданих у заявці чотирьох 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 результуючого профілю якості та відповідних підпрофілів для одного з університетів Великої Британії наведений у табл. 1. Алгоритм формування загального профілю, описаний у документації RAE на офіційному веб-сайті [17], включає процедуру округлення та врахування кожного із підпрофілів відповідно до заданої пропорції.

Для того, щоб мати змогу коректно порівняти результати RAE 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.

Окрім очевидних переваг, які дає профіль якості RAE в термінах престижу чи маркетингу, він визначає розмір фінансування, які буде отримувати від держави кожна із груп установи впродовж декількох років. Розподіл коштів щороку визначається спеціальною формулою, яку оголошує Спеціальна рада Англії з питань фінансування вищої

Табл. 1. Приклад профілю якості та його складових підпрофілів для групи біології Університету Йорка (дані про якісні профілі є публічно доступними на офіційному сайті RAE2008 [17]).

Тип профілю	Відсоток дослідницької активності, що відноситься до 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

освіти<sup>1</sup> – “Higher Education Funding Council for England” (HEFCE). Зокрема, безпосередньо по закінченню процедури RAE у 2008 році використовувалась така формула розподілу фондів:

$$s = p_{4*} + \frac{3}{7}p_{3*} + \frac{1}{7}p_{2*}, \quad (1.1)$$

де  $p_{n*}$  – це значення із загального профіля якості, що припадає на рівень  $n$ , а  $s$  може розглядатися як відносна міра якості наукових досліджень в даній установі та в рамках даної тематики. Саме ця формула далі застосовується нами у роботі для розрахунку відносного показника якості, лише з використанням значень із підпрофіля “Outputs”  $p'_{n*}$  замість значень із загального профіля якості. Використовуючи термінологію попередніх робіт [19, 20], цей відносний показник якості  $s$  ми надалі будемо називати просто *якістю* групи. Відповідно, абсолютний показник якості позначимо великою літерою  $S$  та називатимемо *силою* групи:

$$S = sN, \quad (1.2)$$

$N$  – кількість членів групи.

Нині RAE та схожі процедури оцінювання ефективності наукової діяльності у інших країнах, базуючись на незалежному експертному оцінюванні, вважаються найбільш надійними. Очевидно, їм властиві усі недоліки рецензування, згадані вище. Окрім того, масштабна

<sup>1</sup>У Великій Британії немає чіткого поділу на навчальні та дослідницькі заклади – за рідкісними винятками, усі наукові дослідження виконуються у вищих навчальних закладах.

підготовка та проведення такого роду міроприємства, самі по собі впливають на нормальну роботу наукових установ, тим самим створюючи результат. Такий феномен відомий під назвою закон Гудхарта (“Goodhart’s law”), що у спрощеному формулюванні говорить про те, що коли якась міра сама стає метою, то 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: в певній мірі 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, ці експерти залучаються саме з наукової сфери, тобто значну частку свого робочого часу вони повинні присвячувати додатковим обов’язкам. Враховуючи усе це, періодично з’являються дослідження, що порівнюють рейтинги наукових груп за версією RAE та рейтинги, що базуються на різноманітних наукометричних індикаторах на основі даних про цитування.

## 2. 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 використовувалися значення так званого показника нормалізованого впливу цитувань, normalised citation impact (NCI), що були надані нам компанією Evidence для кожної наукової групи. Значення NCI, яке далі в роботі позначимо малою літерою  $i$ , є відносним. Відповідно, можна розрахувати абсолютне значення ефективності:

$$I = iN, \quad (2.1)$$

$N$  – кількість членів групи. Оскільки наукові цитування лише опосередковано свідчать про якість наукових результатів, а в більшій мірі – про їх вплив на наступні дослідження, то при оперуванні індикаторами на основі цитованості будемо вживати 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 у цитованості різних галузей науки, різних видань чи навіть різних проміжків часу. Загальна кількість одержаних цитувань кожної статті ділиться на середню кількість цитувань для даного видання і в даний рік. В процесі розрахунку NCI для колективу науковців нормовані значення цитованості для кожної статті, включеної до заявки в RAЕ2008, сумуються та усереднюються на кількість членів групи.

### 3. Постановка задачі та результати

Наша мета – порівняти між собою відносні показники  $s$  та  $i$ , що характеризують *якість* та *середній вплив* наукових груп, а також абсолютні показники  $S$  та  $I$ , що свідчать про *силу* та *загальний вплив* колективів в цілому. Як буде показано нижче, на рівні абсолютних показників кореляція є достатньо сильною, тоді як значення відносних показників, а також рейтинги установ, що базуються на цих відносних показниках, досить сильно відрізняються. Ці висновки, в основному, зберігаються для різних дисциплін та не залежать від розмірів групи. При цьому під розміром групи мається на увазі її належність до категорії великої, середньої чи малої. Проблема значення розміру для наукового колективу детальніше досліджувалася у роботах [19, 20], де вперше було поставлене питання про значення розміру наукового колективу для його ефективності. Виявилось, існує 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 стає достатньо великою, перевищивши певне значення  $N_c$ , 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 до її *сили* лише при певних значеннях  $N_k < N < N_c$ . Таким чином, було визначено три категорії наукових груп: малі (при  $N \leq N_k$ ), середні ( $N_k \leq N \leq N_c$ ) та великі ( $N \geq N_c$ ). Порогові значення змінюються в залежності 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 найбільший приріст ефективності дає збільшення кількості членів середньої наукової групи [19, 20]. Зважаючи на можливі відмінності для груп різних розмірів, у нашій роботі окремо виділяються великі групи та сукупність малих і середніх (розглядаємо їх разом, тому що, як правило, великих груп є набагато більше

ніж малих та середніх). Потрібно зауважити, що для різних галузей науки різним є і співвідношення груп різного розміру, хоча майже у всіх випадках проаналізованих дисциплін найбільша кількість саме великих колективів.

Якби показник *середнього впливу* наукової групи, розрахований на основі аналізу цитувань, виявлявся тотожним значенню *якості*, що визначається експертами, то можна було б сміливо заявляти про можливість заміщення трудоємкої процедури рецензування відповідним індексом на основі цитувань. В такому разі кореляція між значеннями  $s$  та  $i$  була б строго лінійною. Для перевірки цієї гіпотези ми обрали три дисципліни: хімію та соціологію, що, відповідно, представляють точні та суспільні науки, а також – географію, що поєднує у собі природничий та 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).

На Рис. 1 зображено відповідні графіки кореляцій  $s$  та  $i$  для обраних дисциплін із врахуванням розміру наукових груп. 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 є сильно розкиданими. Найпростішим методом перевірки якості лінійної апроксимації є розрахунок коефіцієнта Пірсона: чим його абсолютне значення є ближчим до 1, тим краще точки “лягають на пряму”. Розраховані значення коефіцієнта Пірсона, наведені у табл. 2, свідчать про, як правило, 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 та середніх груп.

Побудувавши рейтинг установ на основі  $s$ , ми побачимо, що він сильно відрізняється від аналогічного рейтингу на базі  $i$ . Використаємо спеціальний коефіцієнт Спірмена, що дає змогу порівняти рейтинги між собою. Так само, як у випадку коефіцієнта Пірсона, близьке до 1 абсолютне значення коефіцієнта Спірмена говорить про максимальну подібність між двома рейтингами. Розраховані значення свідчать про те, що рейтинги на основі  $s$  та  $i$  подібні лише на

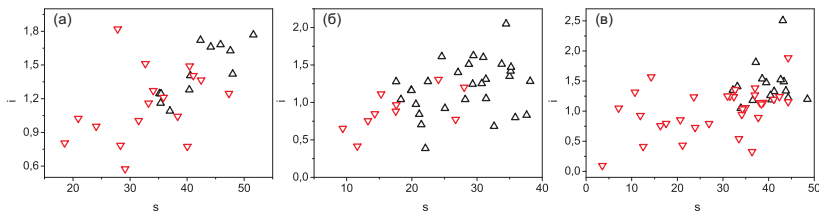


Рис. 1. Залежність між відносними показниками  $s$  та  $i$  для наукових груп в галузі (а) хімії, (б) соціології та (в) географії. Трикутниками донизу показано результати для малих та середніх груп, трикутниками доверху – для великих груп [15].

62% у галузі хімії та близько на 47% у випадках соціології та географії (див. табл. 2). Такі результати суперечать деяким дослідженням, що свідчать про високий рівень кореляції між рейтингами на основі даних RAE та на основі підрахунку цитувань. Наприклад, були знайдені відносно високі значення коефіцієнта Спірмена для деяких специфічних дисциплін (з точки зору методології оцінювання): 0.80 для музики [21] та 0.81 для археології [22].

Табл. 2. Значення коефіцієнтів кореляції між відносними показниками  $s$  та  $i$ , відповідно, якості та середнього впливу наукових колективів Великої Британії для різних дисциплін.

Дисципліна	Коефіцієнт лінійної кореляції Пірсона			Коефіцієнт Спірмена для порівняння рейтингів
	Для усіх груп	Для великих груп	Для малих та середніх груп	
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

У нашому дослідженні можна було очікувати побачити добру кореляцію між рейтингами з огляду на використані більш тонкі методики: можливість врахування окремих підпрофілів, що стало можливим лише для результатів RAE у 2008 році та продуманий алгоритм нормування при розрахунку значень NCI, що є перевагою у

порівнянні із простим підрахунком цитувань. Проте виявилось, що навіть нетривіальна процедура даних про цитування не відтворює результат рейтингування наукових груп на основі експертних оцінок. Тому нормалізований показник впливу на основі цитувань не можна назвати альтернативою рецензування.

Тепер порівняємо абсолютні показники: *загальний вплив* наукової групи  $I$  та її *силу*  $S$ . Знову ж таки, якщо б результат рецензування повністю співпадав із наукометричними індикаторами на основі цитувань, то ці дві величини були б лінійно взаємозалежними. На Рис. 2 показано, як насправді співвідносяться  $S$  та  $I$  для обраних трьох дисциплін. Бачимо, що перемноживши відносні значення  $s$  та  $i$  на розмір групи  $N$  [формули (1.2) та (2.1)], ми одержали досить добре скорельовані абсолютні показники. В усіх трьох випадках чітко видно, що значення дещо розкидані, проте зосереджені в околі прямої лінії. Це є наслідком того, що показники ефективності груп ніби одержують вагу, пропорційну до кількості членів групи: показники великих груп (як статистично більш надійні) тепер мають більше значення, ніж показники малих груп. У Табл. 3 наведені відповідні значення коефіцієнта лінійної кореляції Пірсона, що також вказують на досить сильну лінійну скорельованість величин  $S$  та  $I$ . З огляду на те, що абсолютні показники не резонно використовувати для побудови рейтингів, немає змісту рахувати і коефіцієнти Спірмена.

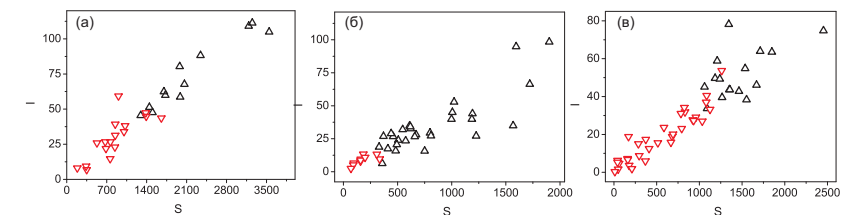


Рис. 2. Залежність між абсолютними показниками  $I$  та  $S$  для наукових груп в галузі (а) хімії, (б) соціології та (в) географії. Трикутниками донизу показано результати для малих та середніх груп, трикутниками доверху – для великих груп [15].

Нормовані або ж відносні показники ефективності наукових груп дають змогу порівняти середню результативність груп різного розміру. Скажімо, за результатами RAE 2008 приблизно однаковий рейтинг одержали групи з історії двох різних установ: Відкритого університету (Open University) та Університету Ґламоргана (University of Glamorgan) при тому, що розміри цих наукових колективів відрі-

Табл. 3. Значення коефіцієнтів кореляції між абсолютними показниками  $S$  та  $I$ , відповідно, *сили* та *загального впливу* наукових колективів Великої Британії для різних дисциплін.

Дисципліна	Коефіцієнт лінійної кореляції Пірсона		
	Для усіх груп	Для великих груп	Для малих і середніх груп
Хімія	0.96	0.96	0.79
Соціологія	0.88	0.82	0.73
Географія	0.92	0.56	0.93

зняються більше, ніж втричі. 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 у деяких випадках для економії часу та зусиль замість процедури експертного оцінювання можна було б використати наукометричні показники на основі аналізу цитувань. У випадку хімії та соціології такий висновок навіть більш справедливий для великих груп (див. Табл. 3), що зрозуміло на інтуїтивному рівні: більша статистика – менші відхилення і надійніший результат. Про причини сильнішої кореляції для малих та середніх наукових груп в галузі географії можна лише припускати. Зокрема, важливою відмінністю географії в контексті нашого дослідження є її сильно 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 сильну кореляцію між показниками *сили*  $S$  та *загального впливу*  $I$  наукових груп ми розрахували також коефіцієнти 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 усіх дисциплін. Висновок про сильну кореляцію між абсолютними значеннями  $S$  та  $I$  також підтверджується в усіх випадках. Зауважимо, що найсильніша кореляція (з коефіцієнтом Пірсона вище 0.90) спостерігається для дисциплін, ближчих до точної науки (або ж до *hard sciences*, за однією з класифікацій, прийнятих в англомовній 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 (або так званих *soft sciences*) – історії та соціології. Відповідні значення коефіцієнтів кореляції для сильно міждисциплінарної географії та прикладної інженерії знаходяться десь посередині, в околі 0.92.

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

Дисципліна	Коефіцієнт лінійної кореляції 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

#### 4. Короткі висновки

Підсумовуючи, можна стверджувати, що на сьогодні немає стовідсоткової альтернативи процедурі експертного оцінювання у сфері науки. Незважаючи на усі недоліки, незалежне рецензування залишається найкращим гарантом якості наукової роботи. Проте не виключається 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 із цитуванням, – все це зумовлює високий рівень похибки при намаганні оцінити наукову працю індивідуума чи цілого колективу. Тому одержані таким чином цифри слабо корелюють із оцінками незалежних експертів. Проте якщо оцінювати загальну ефективність усієї наукової групи, то статистична похибка зменшується тим сильніше, чим більшим є розмір групи. Тому такі зважені оцінки на основі тих же наукометричних індикаторів дають змогу з високою ймовірністю передбачити експертну оцінку загальної сили групи. Ці висновки є справедливими для різних галузей знання, хоча найкраще виконуються для точних наук. Отже, можна сформулювати дві головні тези:



- Оцінки рецензентів та індикатори на основі аналізу цитувань, використані для оцінки середньої ефективності наукових груп, корелюють слабо. Тому не можна застосовувати наукометричні індикатори для формування рейтингів наукових колективів чи установ без врахування думки експертів.
- Добре корелюють між собою оцінки рецензентів та індикатори на основі аналізу цитувань, використані для оцінки загальної ефективності наукових груп. Отже, наукометричні індикатори з певною долею похибки можуть бути використані для виявлення сили або абсолютної ефективності наукових груп.

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

## Література

1. Основы науковедения / [ред. Н. Стефанов, Н. Яхиел, Я. Фаркаш, Г. Кребер, И. Малецкий, С. Микулинский, Р. Рихта]. – М. : Наука, 1985. – 431 с.
2. Жмудь Л. Я. Зарождение истории науки в Античности / Л. Я. Жмудь. – Санкт-Петербург : Изд. Русского христианского гуманитарного института, 2002. – 424 с. – ISBN 5-88812-172-X.
3. Добров Г. М. Наука о науке / Г. М. Добров. – К. : Наукова думка, 1989. – 304 с. – ISBN 5-12-001112-8.
4. Butler D., University rankings smarten up, Nature, 2010, 464, 16-17.
5. Williams R., de Rassenfosse G., Jensen P., Marginson S., U21 Ranking of National Higher Education Systems, Report of the project sponsored by Universitas 21, University of Melbourne, 2012.
6. Bibliometric evaluation and international benchmarking of the UKs physics research, Summary report prepared for the Institute of Physics by Evidence, Thomson Reuters, 2012.
7. Simkin M. V. Stochastic modeling of citation slips / M. V. Simkin, V. P. Roychowdhury // Scientometrics. – 2005. – Vol. 62, № 3. – P. 367-384.
8. Simkin M. V. Read before you cite! [Електронний ресурс] / M. V. Simkin, V. P. Roychowdhury // Ithaca. : Cornell University Library, 2002. – 4 p. – (Preprint / arXiv.org; arXiv: cond-mat/0212043v1).
9. van Raan A.F.J., Fatal attraction: Conceptual and methodological problems in the ranking of universities by bibliometric methods, Scientometrics, 2005, 62, No. 1, 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, 6, e18521.
11. Bornmann L., The Hawthorne effect in journal peer review, Scientometrics, 2012, 91, 857-862.
12. Nature Editorial (2010). Nature, and metrics special. <http://www.nature.com/metrics>. Востанне перевірено доступність у квітні 2012 р. (465, p. 845).
13. Warner J., Citation Analysis and Research Assessment in the United Kingdom, B. Am. Soc. Inform. Sci. Tech., 2003, 30, Iss. 1, 26-27.
14. De Bellis, N. (2009). Bibliometrics and citation analysis: from the science citation index to cybermetrics. USA: The Scarecrow Press, Inc., Lanham, Maryland, Toronto, Plymouth (UK).
15. Mryglod O., Kenna R., Holovatch Yu., Berche B. Absolute and specific measures of research group excellence // Scientometrics. – 2013. – Vol. 95, Issue 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, (accepted for publication), DOI: 10.1007/s11192-013-1058-9.
17. The official web-page of the RAE 2008. <http://www.rae.ac.uk/>. Востанне перевірено доступність 18 березня 2013 р.
18. The official web-page of Thomson Reuters Research Analytic: <http://www.evidence.co.uk>. Востанне перевірено доступність 18 березня 2013 р.
19. Kenna, R., Berche, B. (2010). Critical mass and the dependency of research quality on group size. Scientometrics, 86(2), 527-540.
20. Kenna, R., Berche, B. (2011a). Critical masses for academic research groups and consequences for higher education research policy and management. Higher Education Management and Policy, 23(3), 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, 13, No. 2.
22. Norris M., Oppenheim Ch., Citation counts and the Research Assessment Exercise. V Archaeology and the 2001 RAE, Journal of Documentation, 2003, 59, No. 6, 709-730.

# CONDENSED MATTER PHYSICS

The journal **Condensed Matter Physics** is founded in 1993 and published by Institute for Condensed Matter Physics of the National Academy of Sciences of Ukraine.

**AIMS AND SCOPE:** The journal **Condensed Matter Physics** contains research and review articles in the field of statistical mechanics and condensed matter theory. The main attention is paid to physics of solid, liquid and amorphous systems, phase equilibria and phase transitions, thermal, structural, electric, magnetic and optical properties of condensed matter. Condensed Matter Physics is published quarterly.

**ABSTRACTED/INDEXED IN:** Chemical Abstract Service, Current Contents/Physical, Chemical&Earth Sciences; ISI Science Citation Index-Expanded, ISI Alerting Services; INSPEC; "Referatyvnyj Zhurnal"; "Dzherelo".

**EDITOR IN CHIEF:** Ihor Yukhnovskii.

**EDITORIAL BOARD:** T. Arimitsu, *Tsukuba*; J.-P. Badiali, *Paris*; B. Berche, *Nancy*; T. Bryk (Associate Editor), *Lviv*; J.-M. Caillol, *Orsay*; C. von Ferber, *Coventry*; R. Folk, *Linz*; L.E. Gonzalez, *Valladolid*; D. Henderson, *Provo*; F. Hirata, *Okazaki*; Yu. Holovatch (Associate Editor), *Lviv*; M. Holovko (Associate Editor), *Lviv*; O. Ivankiv (Managing Editor), *Lviv*; Ja. Ilnytskyi (Assistant Editor), *Lviv*; N. Jakse, *Grenoble*; W. Janke, *Leipzig*; J. Jedrzejewski, *Wroclaw*; Yu. Kalyuzhnyi, *Lviv*; R. Kenna, *Coventry*; M. Korynevskii, *Lviv*; Yu. Kozitsky, *Lublin*; M. Kozlovskii, *Lviv*; O. Lavrentovich, *Kent*; M. Lebovka, *Kyiv*; R. Lemanski, *Wroclaw*; R. Levitskii, *Lviv*; V. Loktev, *Kyiv*; E. Lomba, *Madrid*; O. Makhanets, *Chernivtsi*; V. Morozov, *Moscow*; I. Mryglod (Associate Editor), *Lviv*; O. Patsahan (Assistant Editor), *Lviv*; O. Pizio, *Mexico*; N. Plakida, *Dubna*; G. Ruocco, *Rome*; A. Seitsonen, *Zürich*; S. Sharapov, *Kyiv*; Ya. Shchur, *Lviv*; A. Shvaika (Associate Editor), *Lviv*; S. Sokolowski, *Lublin*; I. Stasyuk (Associate Editor), *Lviv*; J. Strečka, *Košice*; S. Thurner, *Vienna*; M. Tokarchuk, *Lviv*; I. Vakarchuk, *Lviv*; V. Vlachy, *Ljubljana*; A. Zagorodny, *Kyiv*

## CONTACT INFORMATION:

Institute for Condensed Matter Physics  
of the National Academy of Sciences of Ukraine  
1 Svientsitskii Str., 79011 Lviv, Ukraine  
Tel: +38(032)2761978; Fax: +38(032)2761158  
E-mail: cmp@icmp.lviv.ua <http://www.icmp.lviv.ua>