

Оцінки результатів наукової діяльності: традиційні підходи та нові виклики*

У роботі розглянуто основні підходи до оцінки результатів наукової діяльності на прикладі провідних країн світу, зокрема США. Показано, що найбільш популярним інструментом залишається використання експертної процедури. Крім того, у роботі аналізуються нові проблеми, що виникли в останні роки при оцінюванні наукових результатів, та запропоновано шляхи їх вирішення.

Постановка проблеми. Оцінка результатів наукової діяльності на різних рівнях управління та в різних галузях економіки, попри зусилля багатьох фахівців, залишається значною проблемою, особливо, коли йдеться про фінансування фундаментальних досліджень. Для такої оцінки використовуються великий арсенал спеціалізованих методів і індикаторів, але можливості їх універсального застосування залишаються обмеженими. Результати майже кожного наукового проекту мають свою специфіку, яку не завжди вдається відобразити за допомогою одних і тих самих показників. У найбільш узагальненому випадку можна вважати, що ступінь досягнення поставленої мети, за умови використання виділених ресурсів, є найважливішим мірилом успішності досліджень і розробок (ДіР). Однак такий підхід не вирішує проблеми порівняння кінцевих результатів різних проектів та не знімає питання про коректність і повноту апріорного визначення мети конкретних науково-дослідних проектів (тобто, ресурси, у принципі, можуть бути витрачені на хибні цілі).

Мета роботи: у контексті аналізу найбільш поширених у світі підходів до оцінки результатів наукової діяль-

ності проаналізувати нові явища, що ускладнюють таку оцінку, та запропонувати відповідні заходи з удосконалення процедур оцінювання наукових результатів в Україні.

Аналіз проведених досліджень та відповідних публікацій. Результати діяльності в економіці порівнюються, як правило, на основі аналізу відповідних кількісних показників, але, на наш погляд, найважливіша проблема в галузі оцінки результатів науково-технічної діяльності полягає в тому, що вимірюється, скоріше, те, що можна виміряти, а не те, що потрібно вимірювати. Недоліки основних кількісних показників результативності, пов'язаних із публікаційною активністю, широко обговорюються в сучасній науковій літературі [1,2]. У якості приклада можемо послатися на ґрунтовний огляд, зроблений Р.Г. Касимовою [3]. Звичайно, що подібні недоліки стосуються й використання індикаторів патентної активності, а у випадку фундаментальних досліджень – навіть ще більшою мірою [4].

Вихід з цієї ситуації різні дослідники шукають у створенні цілісних систем індикаторів, розробці відповідних процедур оцінки та їх постійної адаптації до мінливих умов економічної діяльності. При цьому, однак, у багатьох випадках береться до уваги те, що, наука і технології виконують не тільки економічну функцію. Це знаходить відображення й у процесах вимірювання

*Роботу підготовлено в рамках виконання проекту НАН України „Основні проблеми та напрями інноваційного розвитку України у 2015–2020 рр.”.

результативності ДіР, й у процедурах вибору відповідних індикаторів.

Саме тому для оцінки результативності наукової діяльності широке поширення отримали евристичні методи, зокрема, метод peer review [5]. Це зовсім не означає, що кількісні показники результативності не застосовуються для аналізу результатів діяльності наукових організацій та окремих вчених, але вони виконують допоміжну функцію при формулюванні висновків кваліфікованими експертами. Слід також зазначити, що peer review є, перш за все, найважливішим методом оцінки наукових результатів у галузі фундаментальних досліджень. Ключовою для прийняття рішення є думка експертів про ту чи іншу роботу та її результати. У якості допоміжних індикаторів використовуються різні кількісні показники, але висновок експерта про якісний бік дослідження є визначальним. Саме цей метод використовується для відбору та оцінки наукових проєктів Національним науковим фондом США (ННФ), Національним інститутом охорони здоров'я (НІЗ) та науковими радами у країнах Європи.

Варто більш детально розглянути досвід США як країни, що є найуспішнішою у сфері організації та проведення досліджень і розробок та має надзвичайно великий досвід в удосконаленні систем оцінки результатів ДіР.

В основі державної політики США у сфері науки лежить принцип врахування інтересів різних учасників наукового процесу і «зовнішніх» по відношенню до цього процесу сил. Як правило, за рахунок використання процедур узгодження вдається досягати збалансованих рішень, у тому числі - щодо оцінки результатів досліджень, але «повної» об'єктивності чекати все одно важко.

Ще на початку 1990-х рр. Конгрес США виділив спеціально 3% наукового бюджету для оцінки рівня проведених ДіР у ННФ і НІЗ. При цьому слід зазначити, що приблизно в цей же час у США почалися спроби знайти формулу

для застосування Державного акту щодо оцінки діяльності та результатів (Government Performance and Results Act, 1993) [6], який використовується, зокрема, для аналізу ефективності коштів, що виділяються на різноманітні програми з федерального бюджету. Відповідно до нього, було посилено вимоги до бухгалтерської звітності у сфері проведення ДіР. Раніше для приблизно 20% ДіР, що фінансувалися державою, не було потрібно практично ніякої звітності.

Головне управління з фінансового контролю США (US General Accounting Office) декілька років тому виконало на замовлення Конгресу роботу по узагальненню індикаторів, що використовуються у процесах оцінки ДіР. Відповідно до запропонованої методології, розглядалися чотири основних групи індикаторів:

- поточні індикатори, орієнтовані на аналіз патентної активності (при цьому підкреслено, що підходи до оцінки результатів ДіР у державному і приватному секторах не можуть бути ідентичними);
- бібліометричні індикатори (при цьому зазначено, що частота цитування не є показником рівня новизни досліджень);
- експертна оцінка різних аспектів виконання та результатів проєкту (при цьому зауважено, що сприйняття й оцінка результатів залежать від складу експертів і тому далеко не завжди є об'єктивними);
- комерційний ефект від проведених ДіР – застосовується у приватному секторі (при цьому зазначено, що надзвичайно складно виокремити «чистий ефект» від ДіР, що проводяться, у його відриві від внеску інших видів діяльності).

Тим не менш, усі спроби розробити «чисто» кількісну систему оцінок проведених фундаментальних досліджень у США завершилися невдачею. Було вирішено експертним шляхом визначити відповідність отриманих результатів поставленим цілям. По цей час для

оцінки результатів проектів ННФ і НІЗ зазвичай використовують у першу чергу не кількісні шкали, а якісні судження про те, чи відповідають вони світовому рівню, чи ні. Звіти за проектами ННФ і НІЗ розміщуються в Інтернеті. До проведення експертизи залучаються щорічно тисячі незалежних фахівців. Правда, слід визнати, що оцінку технологічно орієнтованих проектів проводять експерти Національного інституту стандартів разом із представниками приватних компаній, використовуючи приблизно тридцять спеціально відібраних індикаторів [7]. Більше того, у 2012 р. у США було прийнято доповнення до Government Performance and Results Act, але вони стосувалися, знов таки, не конкретних показників, а напрямів удосконалення формування цілей діяльності органів державної влади та контролю над виконанням планів тієї чи іншої державної агенції (оцінка повноти та вчасності досягнення сформульованих цілей за умови виділення відповідних ресурсів), що відповідає за реалізацію наукових проектів [8].

Національний науковий фонд США в останні роки зробив певні зміни у процедурі організації роботи експертів у рамках застосування процедури peer review*. Так, експерти отримують спеціальний грант за виконану ними роботу, а не разову оплату, як практикувалося раніше, по-друге, до роботи в якості експертів залучаються фахівці, які отримали підряд два гранти від державних агенцій на виконання наукових проектів і тому не можуть претендувати на третій. Виконувати свої функції експерти можуть не більше трьох років поспіль з обов'язковою 12-річною перервою після закінчення цього терміну. Такий підхід дозволяє в більшій мірі «об'єктивізувати» процедуру відбору та оцінки результатів проектів. Крім того, Національний

* У вільному перекладі це можна трактувати як оцінювання фахівцем-«метром» у цій дисципліні.

науковий фонд США спростив систему оцінки поданих заявок на отримання грантів. Експертів просять оцінити заявки, виходячи тільки з двох критеріїв: якість наукового змісту проекту і його вплив на суспільство. Раніше, крім двох вищезазначених критеріїв, експерти повинні були оцінювати наукові досягнення фахівця, що підготував пропозиції, і вплив проекту на розвиток інфраструктури ДіР у США.

Цікаво також зазначити, що Національний науковий фонд США повністю перейшов на електронну технологію підготовки заявок через Інтернет. Тепер усі необхідні для підготовки заявки матеріали і навіть консультації по їх заповненню можна отримати безпосередньо, використовуючи веб-сторінку фонду.

Для підвищення ефективності інвестицій федерального уряду у сферу ДіР до оцінки їхніх результатів залучаються, по можливості, максимально ясні інвестиційні критерії. Інвестиційні критерії повинні враховувати, зокрема, як очікувані кінцеві результати, так і прогрес у їхньому досягненні, а також містити показники для оцінки отриманих результатів.

Але в будь-якому разі правомірність використання системи peer review не береться під сумнів. Так, у 2009 р. Національна Академія наук США оприлюднила доповідь, у якій аналізувалися проблеми оцінки фундаментальних досліджень, що проводяться у країні. Експерти дійшли висновку, що фундаментальні результати та значна частка результатів прикладних можуть бути оцінені тільки професіоналами в тій або іншій науковій дисципліні. При цьому бажано, щоб до складу експертних комісій входили й зарубіжні фахівці.

У рамках проведення відповідних реформ та з метою забезпечення порівнянності самих різних видів ДіР (від прикладних досліджень та розробок і демонстраційних програм), були виділені найбільш загальні критерії, що

відбивають три фундаментальних аспекти, властиві будь-якій програмі або проекту ДіР:

1. Відповідність (Relevance) - обґрунтування важливості, можливості і необхідності федерального інвестування програми (проекту).

2. Якість (Quality) – обґрунтування того, як інвестування бюджетних засобів забезпечить найкращу якість ДіР.

3. Результативність (Performance) - обґрунтування ефективного використання витрат.

У цьому контексті рекомендовано звертати особову увагу на такі аспекти наукової діяльності:

1. Наукова продукція і її «наочність» (публікації, патенти і т.п.).

2. Тип фінансування.

3. Гомогенність дослідницької тематики.

Багато проблем у галузі формування науково-технічної політики США та її оцінки на федеральному рівні виникає через те, що деякі формулювання, що використовуються, не є досить чіткими. Так, Конгрес зобов'язав Національний науковий фонд витратити не менше 60% коштів на проведення так званих «стратегічних досліджень». Однак визначення цього типу досліджень не було дано досить чітко. Експерти пов'язують казус, що виник, з бажанням конгресменів посилити роль науково-технічних факторів у протистоянні з конкурентами з Європи і Південно-Східної Азії і прагненням використати ННФ для фінансування не тільки фундаментальних, але і прикладних ДіР. У будь-якому разі, на думку експертів Національного наукового фонду США, система оцінки, що використовується нині федеральними організаціями з управління ДіР, не потребує кардинальних змін.

Водночас в останні роки виникли нові проблеми, пов'язані із специфікою найсучасніших досліджень: використанням унікального обладнання та даних, вартістю проведення експериментів, витратами значного часу на реалі-

зацію дослідницьких програм. Спробуємо узагальнити деякі з цих проблем та навести відповідні приклади.

1. Фальсифікація результатів з метою отримання додаткового фінансування та міжнародного визнання в науковому товаристві. Один з найбільш яскравих епізодів такого роду був зафіксований у 2014 р. Група вчених під керівництвом доктора Х. Обоката з Центру розвитку біотехнологій Інституту фізико-технічних досліджень у Кобе (Японія) повідомила про новий ефективний метод отримання стовбурових клітин у престижному журналі «Нейчр». Проте спроби повторення результатів іншими вченими не увінчалися успіхом. Японська дослідниця була звинувачена в підтасовуванні даних експериментів і під тиском наукової спільноти відкликала свою публікацію [9].

2. Розповсюдження неякісної наукової інформації, що призводить до помилок, які не так легко виявити. Подібний приклад «псевдовідкриття» в галузі астрофізики також відноситься до 2014 р. Американський астроном Джон Ковач у березні повідомив про зроблене ним відкриття слабкого реліктового випромінювання в певному діапазоні радіохвиль [10]. Його дані були доступні в мережі Інтернет, і після низки перевірок виявилось, що насправді ніякого слабкого випромінювання в розглянутому діапазоні немає: підвела точність вимірювань і наявність так званих «шумів». Проте цей випадок віднесено до ненавмисної помилки, хоча, як і в попередньому випадку, мало місце отримання хибного результату.

Обидва випадки, незважаючи на різницю в оцінці науковою громадськістю, об'єднує дещо спільне: результати були незалежно протестовані й обговорені в соціальних мережах ученими різних країн, що дозволило оперативно з'ясувати істину.

3. Зростає важливість специфічних компетенцій, що пов'язано з можливостями використання спеціальних знань та вмінь, які необхідні для отримання

унікальних результатів. Як повідомляє журнал «Економіст», кілька років тому фахівці з американської компанії «Амген» змогли відтворити результати лише п'яти з п'ятдесяти трьох досліджень у галузі боротьби з раком, опублікованих у провідних міжнародних журналах. Учені з корпорації «Байер» були більш успішні: вони змогли повторити результати приблизно чверті досліджень з 67 відібраних для тестування. Зрештою, як випливає з даних, опублікованих Американським національним інститутом охорони здоров'я, в біомедичних дослідженнях дуже складно відтворити приблизно три чверті результатів [11].

4. Дедалі частіше зустрічаються помилки при обробці статистичних даних. Техніка статистичних обчислень постійно вдосконалюється, але далеко не завжди вчені коректно використовують статистичні методи і, особливо, новітні модифікації спеціалізованих статистичних пакетів для комп'ютерної обробки даних.

5. Неготовність до сприйняття негативних результатів та намагання уникнути публікації відповідних статей. Наукові журнали менш охочо друкують статті з негативними результатами, ніж статті з позитивними. Згідно з дослідженням, проведеним Даніелем Фанеллі з Единбурзького університету, з 4600 опублікованих з 1990 р. по 2007 р. робіт, середня частка робіт з негативними результатами коливалася від 10% до 30%. Більш того, загальна частка статей, що містять негативні результати, за період, що розглядався, знизилася з 30% до 14%.

6. Спеціалізація веде до того, що багато помилок у наукових виданнях пропускається. Так, ще в 1998 р. під час «експерименту», що став «класичним», редактор престижного «Британського медичного журналу» Фіона Годлі навмисно розіслала 200 експертам журналу статтю, що містила вісім помилок. Жоден з експертів не виявив усіх помилок. Більшість відзначила не більше однієї помилки, а деякі експерти взагалі

нічого не помітили. У цілому, грубі помилки знаходять менш, ніж у 1% статей з приблизно 1,4 млн, що публікуються щорічно в наукових журналах.

Заради справедливості слід зазначити, що пряма фальсифікація грає у процесах надання невірних результатів не основну роль. Аналіз 21 обстеження, проведеного переважно серед фахівців у галузі біомедичних наук, показав, що тільки 2% респондентів зазначили, що деякі колеги займаються свідомою фальсифікацією результатів, але 28% вказали на використання не цілком надійних методик проведення експериментів та інтерпретації даних.

Зростаюча конкуренція в академічному світі веде до того, що багато вчених відмовляються розкривати важливі деталі своєї роботи. Це унеможливує верифікацію їх результатів і повторення вже проведених експериментів [11].

Як вирішення проблеми, пропонується:

1) Створити спеціальні фонди, які б фінансували повторні дослідження з метою довести правильність (і коректність) отриманих результатів.

2) За прикладом журналу «Нейчр» ввести спеціальний список питань до кожної публікації, що подається до друку, з метою ідентифікації коректності використаної інформації та методів, що застосовувалися при дослідженні.

Висновки. Намагання знайти чіткі кількісні критерії, які б дозволяли однозначно оцінити результати наукової роботи, не призвели до позитивних наслідків. У провідних країнах світу, зокрема у США, експертна процедура залишається основним методом проведення оцінок у науковій сфері, особливо в галузі фундаментальних досліджень. Очевидно, що показники публікаційної активності та рівня цитування, за використання яких наполегливо агітують деякі вітчизняні журналісти та науковці, не повинні бути єдиним критерієм оцінки наукових результатів. У нашій країні центральне місце при цьому повинні відігравати прозорі та логічні процедури із залученням

не тільки українських, але й зарубіжних фахівців. До речі, до практики залучення зовнішніх експертів у подібних випадках вдаються не тільки країни Центральної та Східної Європи, але й Австрія, Швейцарія, Швеція, Норвегія.

На основі лише наявних даних статистики неможливо зробити висновки щодо ефективності використання бюджетних коштів, витрачених на дослідження і розробки, особливо в галузі фундаментальних досліджень. Це можна пояснити трьома причинами:

– по-перше, результати науково-технічної діяльності важко оцінити повною мірою, адже вони «розтягнуті» у часі і можуть бути використані не одними, а різними економічними агентами (компаніями), які до того ж належать до різних секторів економіки.

– по-друге, частину результатів важко визначити кількісними показ-

никами. Практично всі висновки щодо ефективності заходів державної підтримки наукової сфери можна зробити на основі лише «непрямого» аналізу та з певним ступенем вирогідності.

– по-третє (і це вже стосується безпосередньо України), держава не завжди в повному обсязі та в узгоджені терміни фінансує відповідні роботи, що дає виконавцям аргументи для неадекватного виконання взятих на себе зобов'язань.

В останні роки виникли нові проблеми щодо оцінки наукових результатів, пов'язані з ускладненням проведення експериментів та труднощами верифікації результатів. Тут необхідно, напевно, вивчати практику розвинених країн щодо надання додаткових даних та розкриття первинної інформації, що використовувалася при проведенні наукових проєктів.

1. *Barker K.* Strengthening the Impact of R&D Evaluation on Policy Making: Methodological and Organisational Considerations // *Science and Public Policy*. – 1994. – Vol. 21. – N.6. – P. 405–413.

2. *Cozzens S.* Assessing federally-supported academic research in the United States // *Research Evaluation*. – 2000. – Vol.9. – N.1. – P.5–10.

3. *Касимова П.Г.* Наукометрические показатели как один из индикаторов качества научной деятельности // *Науковедение*. – 2002. – №1. – С. 132–143.

4. *Rojo J., Polt W.*, eds. Handbook on the Evaluation of Research and Technology Policy: Concepts, Tools, Indicators. – N.Y. and Cheltenham: Edward Elgar, 2003. – 208 p.

5. *Teich A., Nelson S., Lita S.*, eds. Science and Technology in a Vulnerable World: Supplement to AAAS Science and Technology Policy Yearbook 2003. – Washington DC (USA): AAAS, 2003. – 73 p.

6. *Government Performance Results Act of 1993*. – <http://www.whitehouse.gov/omb/mgmt-gpra/gplaw2m>

7. *NSF Science and Engineering Indicators*. – NSF: Arlington, 2012, vol. 1, 576 p.

8. *Brass C.* Changes to the Government Performance and Results Act (GPRA): Overview of the New Framework of Products and Processes. – CRS Report to the Congress. – Congressional Research Service. Washington. DC, 7-5700 R42379. – 2012. – 36 p.

9. *Servik K.* Researcher Behind Stem Cell Controversy Agrees to Retraction. – <http://news.science-mag.org/asiapacific/2014/05/researcher-behind-stem-cell-controversy-agrees-retraction>.

10. <http://astronomy.fas.harvard.edu/news/john-kovac-assistant-professor-harvards-astronomy-and-physics-departments-has-been>.

11. *Let the light shine in* // *The Economist*. – 2014. – June 14. – Vol. 411. – N.8891. – P. 87–88.

Одержано 15.07.2014

И.Ю. Егоров

Оценки результатов научной деятельности: традиционные подходы и новые вызовы

В работе рассмотрены основные подходы к оценке результатов научной деятельности на примере ведущих стран мира, в частности США. Показано, что наиболее популярным инструментом остается использование экспертной процедуры. Кроме того, в работе анализируются новые проблемы, возникшие в последние годы при оценивании научных результатов, и предложены пути их решения.