

## ARTÍCULO ESPECIAL

# Evaluación de la investigación biomédica\*



Jordi Camí

*Institut Municipal d'Investigació Mèdica. Universitat Pompeu Fabra. Barcelona.*

Cuantificar y evaluar la actividad científica es una tarea compleja. Comienza a ser habitual que nos inviten a examinar cifras acerca de los recursos destinados a investigación y desarrollo, cifras que aparecen en memorias institucionales, estudios sectoriales o medios de comunicación. Se trata de datos estadísticos normalizados sobre inversiones o sobre personal dedicado a actividades científicas y técnicas, cifras que permiten compararnos con los países de nuestro entorno; sin embargo, se trata de una información que dice poco acerca de los resultados que se obtienen y de su relevancia.

En los países industrializados, uno de los retos de la próxima década consistirá en evaluar y rendir cuentas de los programas de gasto público en investigación de forma habitual, sistematizada y comparable. Una pretensión dirigida tanto a minimizar las distorsiones de las intervenciones gubernamentales como a optimar sus efectos. Pero todo ejercicio de evaluación es una actividad compleja y generalmente bastante impopular entre la comunidad científica. Ser sujeto de una evaluación siempre inquieta, incluso puede irritar si la metodología es superficial e imperfecta. Aunque los costes de una evaluación exigente sean muy elevados, sus beneficios suelen ser enormemente rentables.

La disposición normalizada de indicadores es complicada, ya que la ciencia es una actividad colectiva y acumulativa, y un mismo indicador puede interpretarse de forma distinta según el área de conocimiento que se estudie. Además, ninguno de los indicadores es en sí mismo determinante, ya que cada uno de ellos muestra solamente una faceta de la realidad, de manera que sólo tienen verdadera utilidad cuando se consideran en conjunto<sup>1</sup>. Por ello no debe extrañar que amplios sectores de la comunidad científica cuestionen ferozmente la metodología para la evaluación de la investigación biomédica, incluso muchos ponen en tela de juicio que esta actividad sea verdaderamente evaluable.

### La Government Performance and Results Act

El detonante más influyente de la actual corriente evaluadora norteamericana procede de una decisión legislativa de 1993, norma cuya aplicación se ha ido extendiendo progresivamente a través de la Oficina de Gestión y Presupuestos de los Estados Unidos y que se conoce por la Government Performance and Results Act. En virtud de esta ley del Congreso norteamericano, todas las instituciones y agencias fe-

derales de financiación y promoción de investigación tienen la obligación de establecer sus objetivos estratégicos y darles la suficiente publicidad, han de mostrar el comportamiento de los principales indicadores de gestión y evaluar el rendimiento de sus infraestructuras, procesos y resultados. Por consiguiente, todas las agencias federales de investigación financiadas públicamente tienen la obligación de definir un plan estratégico quinquenal, un plan de acción anual y un plan de evaluación de los resultados de la investigación; unos planes que deben ser el resultado de un proceso abierto, interactivo y participativo, exigiéndose además que el público sea informado de forma inteligible y verificable acerca de la justificación y los resultados esperados de los programas de investigación.

Fue precisamente esta decisión legislativa la que desencadenó una formidable reflexión en la propia comunidad científica norteamericana acerca de los procedimientos adecuados para la evaluación de los resultados de la investigación científica. En efecto, el Committee on Science, Engineering and Public Policy norteamericano (COSEPUP), un comité conjunto formado por tres instituciones (la National Academy of Sciences, la National Academy of Engineering y el Institute of Medicine), procedió a una amplia consulta entre la comunidad científica y emitió unas conclusiones, a nuestro entender, enormemente relevantes. Al fin y al cabo, fue un encargo dirigido a identificar mejoras en los procedimientos de evaluación de los objetivos y resultados de la investigación, identificar mejoras para incorporar las actividades de investigación en la planificación estratégica y establecer estándares de excelencia en la gestión de los programas de investigación.

En el informe al que nos referimos del COSEPUP<sup>2</sup> destacan fundamentalmente tres grandes conclusiones. La primera y más importante es que la evaluación no sólo es técnicamente posible en investigación aplicada, sino que también es realizable en investigación fundamental. No sólo posible, sino que los expertos consideraron que debe efectuarse de forma regular, para lo que es imprescindible que previamente se hayan establecido los correspondientes planes estratégicos en los que se definan los objetivos específicos. La segunda conclusión es que no pueden utilizarse la misma metodología ni la misma temporalidad para la evaluación de los distintos tipos de investigación según su carácter. Es decir, mientras que la investigación aplicada puede ser evaluada mediante el instrumental convencional utilizado en la evaluación de programas industriales y de servicios públicos, la investigación fundamental debe analizarse teniendo en cuenta su carácter impredecible y su horizonte a más largo plazo. Más concretamente, se concluye que la investigación fundamental debe evaluarse mediante análisis retrospectivos dirigidos a determinar estos tres componentes nucleares: calidad, relevancia y liderazgo. Por último, la tercera conclusión consiste en afirmar que, para el análisis de los tres componentes antes mencionados, el mejor sistema sigue siendo la revisión por homólogos expertos. Un procedimiento de revisión que exige el concurso

\*Presentado en parte en el Simposium «La investigación biomédica en los hospitales», dirigido por el Prof. José María Segovia de Arana. Ministerio de Sanidad y Consumo, Consejo Asesor de Sanidad. Madrid, 4-5 de mayo de 1999.

Correspondencia: Dr. J. Camí.  
Institut Municipal d'Investigació Mèdica.  
Dr. Aiguader, 80. 08003 Barcelona.

Correo electrónico: [jcamí@imim.es](mailto:jcamí@imim.es)

Recibido el 14-5-2001; aceptado para su publicación el 18-7-2001

*Med Clin (Barc)* 2001; 117: 510-513

de los mejores científicos, los más cualificados, y que tiene elevados costes de tiempo y dinero. Aunque es imprescindible que la comunidad científica participe y se sienta involucrada en el proceso, para una buena consecución de estos fines no basta con seleccionar a meros homólogos; de lo que se trata es de que los responsables o administradores del proceso de evaluación sean capaces de seleccionar a los mejores, a los homólogos más capacitados e independientes para cada caso específico. Incluso para la determinación del liderazgo, los propios norteamericanos reconocen la necesidad de que en la evaluación participen expertos extranjeros.

### Acerca de la metodología disponible

Una revisión crítica tanto de los instrumentos como de los criterios de evaluación al uso obliga a situar en su debida posición la utilización exclusiva de indicadores parciales o meramente cuantitativos, una práctica que, por lo menos en nuestro medio, está excesivamente extendida<sup>3</sup>. Además, los responsables de ejercer la evaluación deben asumir que es un proceso altamente costoso. En cualquier caso, es importante escoger bien los instrumentos y medidas, así como utilizarlos eficientemente con el fin de minimizar esfuerzos improductivos. Parafraseando a los autores del informe del COSEPUP antes mencionado «la métrica debe ser simple, sin manipulaciones, y debe potenciar su buen uso».

Entre los distintos métodos de que se dispone y que deben combinarse convenientemente destacan los 6 siguientes: los análisis bibliométricos; las tasas de retorno económico y sanitario de la inversión en cada proyecto; la revisión por pares; los estudios de casos; los análisis retrospectivos y los juicios o análisis comparativos con referencias externas, generalmente internacionales (*benchmarking*). Cada uno de estos métodos exige procedimientos distintos y, aisladamente, tiene tantas ventajas como limitaciones; de ahí la importancia de manejar distintos indicadores simultáneamente. Sin ánimo de ser exhaustivos, se comentan a continuación las características principales de cada uno de ellos. Los análisis bibliométricos (tanto de publicaciones en revistas como de patentes) constituyen el principal método cuantitativo de que se dispone, extraordinariamente útil cuando se utilizan grandes cifras. Mediante la bibliometría pueden obtenerse índices de calidad de grandes programas o áreas de conocimiento. Incluso cuando se trata de análisis temáticos específicos, los indicadores bibliométricos pueden estar estrechamente relacionados con los resultados de la revisión por pares<sup>4</sup>. No obstante, la bibliometría no es aplicable a todos los programas o ámbitos del conocimiento, dificulta la comparación entre áreas y países entre sí y puede manipularse artificialmente con cierta facilidad<sup>3</sup>. En este sentido, resultan irritantes algunos análisis superficiales que se observan en nuestro medio, muy en particular aquellos que, ignorando las citas recibidas por las publicaciones, se basan exclusivamente en inferencias del factor de impacto de las revistas de publicación.

La revisión por pares (*peer review*) es el sistema probablemente más utilizado, mejor comprendido y, por ello, también el que recibe el mayor número de críticas. Mediante la revisión por pares los científicos nos hemos dotado de un procedimiento para otorgar el mérito que alcanza un resultado o una propuesta de investigación. Este método permite evaluar la calidad de una investigación e incluso, a veces, anticipar otros componentes como su relevancia (actual o potencial). Aunque es «el menos malo» de los procedimientos disponibles, presenta una gran variabilidad y no sólo exige una adecuada selección de los miembros de los paneles,

sino también un enorme esfuerzo y compromiso de los participantes. Sobre la revisión por pares existe una amplia y permanente bibliografía crítica al respecto. Destáquese quizás el reciente informe liderado por Bruce Alberts, presidente de la norteamericana National Academy of Sciences y dirigido a examinar la organización y función de los procesos de revisión que lleva a cabo el Center for Scientific Review (CSR) de los NIH<sup>5</sup>.

En el mencionado informe, dirigido a actualizar las actividades del CSR, sobresale el interés por mejorar los actuales instrumentos de evaluación para obtener sistemáticamente «una valoración suficiente del proyecto –de su significación, relevancia, metodología, innovación, competencia investigadora y entorno– que permita pronosticar cuál será su impacto sustancial en el progreso de la comprensión de los sistemas biológicos, en la mejora del control de las enfermedades o en el mantenimiento o mejora de la salud»<sup>5</sup>. Entre otras conclusiones preliminares, se propone la necesidad de reconocer e incentivar la capacidad de flexibilidad de los equipos y los proyectos para responder al ritmo de los rápidos cambios en los métodos y las oportunidades científicas. También se reconoce la necesidad de combatir el conservadurismo y la consolidación funcionarial, e incentivar la asunción de riesgos en la organización y en los grupos de investigación. Para ello se propone la necesidad de reestructurar las unidades de investigación a fin de que incluyan las distintas dimensiones de la investigación fundamental subyacente en todos los sistemas fisiológicos y patológicos, así como eliminar posiciones monopolistas en el acceso a la financiación o en el uso de recursos críticos en determinadas áreas de investigación. En relación con el proceso de revisión, se destaca de nuevo la necesidad de depurar los elementos de «clientelismo» propios de la revisión por pares, como las redes de «viejas guardias», y dar plena transparencia a los criterios y perfiles de los evaluadores. En definitiva, se pretende aumentar la imparcialidad y el rigor en la identificación de la investigación de mayor calidad, así como incorporar evaluadores con capacidad acreditada para evaluar áreas punta y zonas «frontera» en la investigación de última generación.

Los «estudios de casos» nos ilustran acerca de los efectos de los factores institucionales, organizativos y técnicos que influyen a todo proceso de investigación, y es una metodología capaz de identificar un amplio abanico de beneficios y productos resultantes de una actividad científica concreta. Los «estudios de casos» pueden implicar simultáneamente a varios programas y áreas del conocimiento, pero su materialización es larga y costosa. Por otra parte, los «análisis retrospectivos» son una modalidad avanzada de «estudios de casos» que comportan la reconstrucción histórica de los procesos y productos inherentes a una actividad científica. Por ello están especialmente indicados para los proyectos de investigación básica de larga maduración. Son útiles para la identificación de vínculos entre programas distintos e innovaciones a lo largo de extensos intervalos de inversión en investigación, pero no son adecuados para la evaluación a corto plazo debido al gran intervalo existente entre el momento en que se realiza una investigación básica y aquel en que se identifican sus resultados prácticos. Finalmente, cobra cada vez mayor importancia el denominado *benchmarking*, metodología que se concentra en el análisis comparativo de los hitos alcanzados, generalmente desde una perspectiva internacional. El *benchmarking* sólo puede efectuarse en áreas temáticas o en líneas de investigación concretas, si bien es uno de los mejores procedimientos para comparar programas y países entre sí. Como ejemplo de referencia puede citarse el estudio que efectuó

la National Science Foundation (NSF) norteamericana acerca de la investigación básica en inmunología.

### **Reflexiones acerca de los retornos (económicos) de la inversión en I+D biomédica**

La tasación del retorno económico, una métrica cuantitativa muy desarrollada desde el campo de la economía aplicada, permite determinar los beneficios económicos de la investigación, aunque de forma limitada y sólo en algunos casos particulares de la ciencia. Esta metodología sólo es útil para una determinada clase de programas o áreas temáticas, ya que no cuantifica otro tipo de beneficios, como los sociales. Además, el tiempo que se requiere para observar los retornos puede ser muy largo (decenas de años). Para el caso de este tipo de análisis en la investigación biomédica se nos ocurre un conjunto de comentarios, nada ajenos al día a día de los investigadores clínicos. En primer lugar, debe aceptarse que, en biomedicina y ciencias de la salud, a menudo es muy difícil distinguir entre las contribuciones procedentes de la investigación fundamental y las de la aplicada (utilizo expresamente esta terminología convencido de que la investigación clínica puede ser tanto fundamental como aplicada, y deseando que deje de considerarse la investigación clínica como lo opuesto a investigación básica, si bien distinta de la investigación que se efectúa en el laboratorio). Aunque todo desarrollo o aplicación generalmente viene precedido de una contribución seminal, la realidad nos dice que tal secuencia no es un axioma, existiendo circunvalaciones entre las contribuciones fundamentales y las aplicadas que, en definitiva, justifican las dificultades existentes en la planificación estratégica de los avances científicos.

En segundo lugar, en biomedicina y ciencias de la salud también es difícil discernir entre las contribuciones procedentes de la inversión pública y las que provienen de la inversión privada. El paradigma, a mi entender, lo constituye el desarrollo de nuevos medicamentos en España, un proceso que, en determinadas etapas, requiere la participación y el consumo de recursos de la asistencia sanitaria pública, unos recursos que a veces se utilizan de forma oculta e indirecta y que nunca han sido debidamente cuantificados. En tercer lugar, muchas innovaciones en clínica y en salud proceden de esfuerzos científicos ajenos a los de la investigación biomédica. Un ejemplo paradigmático es la tecnología láser; ¿quién hubiera pensado que, tras los trabajos teóricos y experimentales de la interacción luz-átomos, se desarrollaría la fenomenal y diversificada tecnología láser, que ya se utiliza habitualmente en nuestros centros sanitarios? En cuarto lugar, existen aún enormes dificultades para asignar innovaciones biomédicas a mejoras específicas en la salud, sobre todo por la importancia que, en paralelo, viene reconociéndose a los factores vinculados a los estilos de vida y el ambiente, así como por la relevancia de otras intervenciones en la esfera de la salud pública. En quinto lugar, cuando se trata de evaluar mejoras en la salud de los ciudadanos, persiste aún un gran número de retos conceptuales. Éste sería el caso, por ejemplo, de aquellas innovaciones que prolongan la vida, ya que generalmente comportan un aumento de las transferencias económicas de generaciones productivas, las jóvenes, a generaciones menos productivas, las ancianas.

### **La evaluación como instrumento para el establecimiento de prioridades**

La necesidad de establecer prioridades en ciencia es inexorable, puesto que no existen recursos para todos los ob-

jetivos posibles y, no menos importante, porque los esfuerzos de la administración pública en el fomento de la I+D son el resultado de decisiones fundamentalmente políticas. No obstante, en nuestro medio el establecimiento de prioridades casi siempre es un mero y flamante ejercicio retórico, generalmente de carácter incrementalista, y son excepcionales las tomas de decisiones de apoyo selectivo a objetivos específicos<sup>6</sup>. Si se establece que un determinado problema de salud u objetivo científico es prioritario, entonces hay que valorar si es posible afrontarlo con éxito, no sólo en el tiempo y con los recursos disponibles, sino con los efectivos científicos de que se dispone. Por ello es inútil pretender establecer prioridades si no se cuenta con mecanismos adecuados de evaluación implantados de forma más o menos sistemática. Sin evaluación, la toma de decisiones políticas es un ejercicio de improvisación que sólo favorece intereses creados y actitudes conservadoras. Nos queda, pues, mucho camino por recorrer, ya que las agencias de fomento de la I+D de nuestro país aún no han incorporado la evaluación como herramienta de trabajo en la planificación estratégica.

Con más motivo resulta extraordinariamente incómodo tener que afrontar pretensiones de algunos dirigentes sanitarios empeñados en «territorializar» la actividad de investigación biomédica y en ciencias de la salud. Basta acudir a sencillas informaciones bibliométricas, como el aumento progresivo de la coautoría, para comprobar la progresiva internacionalización del origen de los avances en biomedicina y salud. Desde esta perspectiva, termina siendo cansino el discurso de algunos dirigentes políticos y gestores de la administración empeñados en establecer prioridades de orden territorial, o incluso que se cuestione el porqué no se investiga al servicio exclusivo de los intereses del propio hospital. Ante este tipo de conflictos, la mejor pedagogía de «urgencias» consiste en señalar la sustancial diferencia (aunque no siempre en métodos y técnicas) de lo que compete al «servicio de estudios» de la dirección del centro con respecto a los objetivos y proyección de la investigación científica. En nuestro sistema sanitario aumenta progresivamente la madurez de nuestras organizaciones en cuanto a estos temas<sup>7</sup> y, cuando menos, es de justicia reconocer que los primeros esfuerzos de las agencias de fomento en cuanto a la evaluación biomédica en España fueron una realidad<sup>8</sup>, si bien ignoramos si existe continuidad alguna al respecto.

### **Conclusión**

Los científicos deberían recibir la evaluación de sus tareas como una oportunidad, no como una amenaza. Los responsables y gestores de las agencias de fomento y coordinación de la investigación de nuestro medio deberían descubrir la importancia de la evaluación de resultados como el instrumento más eficaz para sus políticas y como el único medio para rendir cuentas ante los contribuyentes y la sociedad en general. Esta toma de conciencia no debe soslayar que la evaluación de la investigación científica es un ejercicio complejo, bastante caro y que debe ponerse en manos de personas preparadas. La evaluación por parte de aficionados sólo conduce a abusos e injusticias. Precisamente porque lo más prudente es lo que requiere mayor esfuerzo, es decir, el impulso decidido de las evaluaciones basadas en la revisión por pares, mediante paneles de expertos externos e independientes y, a ser posible, compuestos por científicos de talla internacional. Si se piensa exclusivamente en términos de años fiscales o en periodos electorales, si se calcula exclusivamente sobre la base del «corto plazo», tardaremos

en ser obsequiados con evaluaciones rigurosas y útiles. Mientras tanto, debería minimizarse aquello que es más sencillo y, a su vez, más imprudente e injusto. Nos referimos al consumo diletante de indicadores bibliométricos, indicadores extraordinariamente útiles e informativos, pero a riesgo de que alcancen un desprestigio y desconfianza que no será fácil de corregir.

#### REFERENCIAS BIBLIOGRÁFICAS

1. Barré R, Papon P. Indicadores: finalidad y límites. En: UNESCO, editores. Informe mundial sobre la ciencia. Madrid: Santillana, 1993; 134-136.
2. COSEPUP. Research and the Government Performance and Results Act. National Academic Press, 1999.  
<http://www.nap.edu/readingroom/books/grpa>.
3. Camí J. Impactología: diagnóstico y tratamiento. Med Clin (Barc) 1997; 109: 515-524.
4. Rinia EJ, Van Leeuwen ThN, Van Vuren HG, Van Raan AFJ. Comparative analysis of a set of bibliometric indicators and central peer review criteria. Evaluation of condensed matter physics in the Netherlands. Research Policy 1998, 27: 95-107.
5. Recommendations for change at the Center for Scientific Review. Phase I: Report panel on scientific boundaries for review, 1999.  
<http://www.csr.nih.gov/bioopp/intro.htm>.
6. Camí J. Priorización de la investigación biomédica: implicaciones para la investigación en salud pública y servicios sanitarios. Gac Sanitaria 2000; 14: 327-329.
7. Camí J. La evaluación como síntoma de buena salud. Med Clin (Barc) 1999; 112: 218-219.
8. Muñoz Ruiz E, Sanz Menéndez L, Espinosa de los Monteros J, Camí J. Evaluación del FIS como programa de I + D. En: Ricoy JR, Guasch MF, Clavería L, editores. Fondo de Investigación Sanitaria (1988-1995). Una aproximación al análisis de un Programa de I+D. Madrid: Ministerio de Sanidad y Consumo, 1996: 61-75.