



**MONOGRAFÍAS DE POLÍTICA Y GESTIÓN**

# El rol de la Investigación en Efectividad Comparada (CER)

JOSÉ FRANCISCO HORGA





**MONOGRAFÍAS DE POLÍTICA Y GESTIÓN**

# El rol de la Investigación en Efectividad Comparada (CER)

JOSÉ FRANCISCO HORGA

Cualquier forma de reproducción, distribución, comunicación pública o transformación de esta obra solo puede ser realizada con la autorización de sus titulares, salvo excepción prevista por la ley. Diríjase a CEDRO (Centro Español de Derechos Reprográficos, [www.cedro.org](http://www.cedro.org)) si necesita fotocopiar o escanear algún fragmento de esta obra

© 2015 Ergon.  
C/ Arboleda, 1. 28221 Majadahonda (Madrid)

ISBN: 978-84-16270-64-4  
Depósito Legal: M-32288-2015  
Impreso en España

## Autor

---

JOSÉ FRANCISCO HORGA DE LA PARTE  
*Jefe de Sección de Farmacología Clínica.*  
*Hospital General Universitario de Alicante.*  
*Departamento de Farmacología, Pediatría y Química Orgánica.*  
*Universidad Miguel Hernández. Elche, Alicante.*



# Índice

---

|                                                                                                                                       |           |
|---------------------------------------------------------------------------------------------------------------------------------------|-----------|
| PRÓLOGO .....                                                                                                                         | 7         |
| PRÓLOGO .....                                                                                                                         | 9         |
| PRINCIPALES ABREVIACIONES.....                                                                                                        | 11        |
| <b>I. ANTECEDENTES. CONCEPTOS Y CONTEXTOS.....</b>                                                                                    | <b>13</b> |
| • Como método de generación de información, sólo son válidos los ensayos clínicos: El momento y el lugar de la decisión. ....         | 14        |
| • La justificación de la decisión: Las virtudes de los ensayos clínicos controlados y aleatorizados.....                              | 16        |
| – Objetivo primario.....                                                                                                              | 17        |
| – Objetivos secundarios .....                                                                                                         | 18        |
| – Sesgos y errores de interpretación .....                                                                                            | 19        |
| – El ensayo controlado .....                                                                                                          | 19        |
| – Aleatorización.....                                                                                                                 | 20        |
| <b>II. APROXIMACIONES A LA EFECTIVIDAD COMPARADA SIN DATOS DE ESTUDIOS DE EFECTIVIDAD COMPARADA. APORTACIONES Y LIMITACIONES.....</b> | <b>23</b> |
| – Resultado del ensayo: eficacia comparada vs respuesta al tratamiento .....                                                          | 23        |
| – Población en estudio. Validez externa de los datos del estudio. Extrapolación y generalización .....                                | 25        |
| – Análisis de subgrupos .....                                                                                                         | 26        |
| • ¿Un paso adelante? Medicina basada en la evidencia, metaanálisis, revisiones sistemáticas y guías terapéuticas .....                | 29        |
| – Medicina basada en pruebas (Evidence-Based Medicine; MBE) .....                                                                     | 30        |
| – Las Guías Terapéuticas basadas en las pruebas (Evidence-Based Guidelines; GTE) y el nivel de evidencia .....                        | 34        |
| <b>III. EL ROL DE LA CER: DECISIONES, IMPLICACIONES Y APORTACIONES... </b>                                                            | <b>41</b> |
| • Conceptos relacionados con la efectividad clínica y la CER.....                                                                     | 42        |
| • CER.....                                                                                                                            | 44        |
| • Definición de CER.....                                                                                                              | 45        |
| • Estructura y aspectos organizativos de CER en USA .....                                                                             | 46        |
| • Programas tipo CER en otros países. Diferencias con el programa IEC norteamericano.....                                             | 47        |
| – CER y la Unión Europea .....                                                                                                        | 48        |
| – Los análisis de efectividad comparada en España .....                                                                               | 49        |

|                                                                                                                             |           |
|-----------------------------------------------------------------------------------------------------------------------------|-----------|
| • Las implicaciones de CER en la investigación clínica.                                                                     |           |
| La metodología de CER .....                                                                                                 | <b>50</b> |
| – Ensayos clínicos aleatorizados y CER .....                                                                                | <b>52</b> |
| – Cambios en el modelo de estudio: los ensayos clínicos<br>aleatorizados pragmáticos (ECAP) .....                           | <b>53</b> |
| – Cambios en el procedimiento de análisis de los datos:<br>el papel potencial de la aproximación bayesiana adaptativa... .. | <b>55</b> |
| – Estudios observacionales y CER .....                                                                                      | <b>56</b> |
| – Registros de pacientes.....                                                                                               | <b>59</b> |
| – El cuidado centrado en el paciente en la IEC.....                                                                         | <b>60</b> |
| <br>A MODO DE EPÍLOGO .....                                                                                                 | <b>63</b> |
| <br>BIBLIOGRAFÍA .....                                                                                                      | <b>65</b> |
| <br>SOBRE EL AUTOR .....                                                                                                    | <b>73</b> |

## Prólogo

---

Los sistemas sanitarios de los países desarrollados, como es sabido, se están enfrentando a una serie de problemas comunes: el envejecimiento de la población, el incremento de su esperanza de vida; la introducción, acceso y financiación de la innovación; la imprescindible gestión de la evolución de estos sistemas en una situación económica sin precedentes...

En un escenario desafiante, con estos y muchos otros retos, es esencial la racionalización de la incorporación de las nuevas tecnologías sanitarias y la utilización de la mejor evidencia científica disponible, de forma balanceada y crítica, en el proceso de toma de decisiones. Además, como elemento indispensable, y apenas desarrollado en nuestro medio, la reevaluación de la efectividad de estas nuevas tecnologías a lo largo del tiempo es una cuestión ineludible a medio y largo plazo.

En este sentido el papel de la investigación en efectividad comparada, como herramienta de gestión para el mantenimiento de ciertas tecnologías en la cartera de servicios, se antoja fundamental. Tanto, que su difusión e implementación en el proceso de toma de decisiones diaria, seguramente, será una realidad en un futuro no lejano. Para la consecución de este objetivo es necesario el establecimiento de unas bases sólidas, tanto conceptuales como metodológicas.

La Fundación AstraZeneca, desde sus inicios hace doce años, siempre ha perseguido colaborar en la construcción del mejor entorno sanitario posible para los profesionales y los pacientes en nuestro país. La elaboración de esta monografía sobre El rol de la Investigación en Efectividad Comparada -en estrecha colaboración con la Fundación Gaspar Casal y el equipo de Health Economics & Outcomes Research de AstraZeneca España- aspira de nuevo, como el resto de monografía de la serie sobre política sanitaria, a generar y dinamizar el debate entre los profesionales y decisores sanitarios acerca del papel de esta disciplina y su potencial incorporación a su día a día.

Así, en esta monografía, desde la presentación de los antecedentes y la revisión de la aportación de los ensayos clínicos, se profundiza en: los principios de la investigación de la efectividad comparada, su lugar en el contexto de la investigación clínica y se revisan experiencias en EEUU, Europa y España.

En nuestra aspiración de lograr un sistema sanitario sostenible, seguiremos poniendo todos nuestros recursos al servicio de la investigación, el análisis, la reflexión y el debate para encontrar nuevas fórmulas que benefician a todos los agentes implicados, especialmente a los pacientes.

**Guillermo de Juan**

*Director de Asuntos Corporativos y Legal.  
AstraZeneca Farmacéutica Spain S.A.*



## Prólogo

---

La Investigación de la Efectividad Comparada (CER) ha sido definida como la generación y síntesis de evidencia que compara los beneficios y perjuicios de los métodos alternativos disponibles para prevenir, diagnosticar, tratar y monitorizar una condición clínica o para mejorar la administración del cuidado sanitario. Se trata de construir los pilares sobre los que se pueda sustentar un cuerpo de evidencia científica que ayude a cualquier sistema sanitario a tomar decisiones no sólo buscando el mejor resultado clínico, sino también optimizando la efectividad de los recursos. Por consiguiente, la evidencia generada por la CER debería servir a los consumidores, a los clínicos, a los proveedores y a los planificadores en la toma de decisiones informadas para mejorar la asistencia sanitaria tanto a nivel individual como poblacional.

El presidente Obama consideró que la Investigación en Efectividad Comparada (CER) ayudaría a resolver los problemas del sistema sanitario de los EE.UU. y ha dado los primeros pasos para poner a prueba esta hipótesis. En los dos últimos años, la legislación federal ha puesto en marcha leyes, fundado organismos, formado comités, desarrollado prioridades, y las agencias federales han gastado una enorme cantidad de dinero –esperemos que de forma rápida y sabia para financiar la CER. Se han puesto los medios, y ahora habrá que esperar a los resultados. Bienvenida sea la idea, pudiera cundir el ejemplo y que algún líder sanitario local apostara por ello. El futuro de la CER es brillante. No existe razón alguna para esperar que en el futuro próximo estas estrategias sirvan para atenuar el incesante aumento de los costes sanitarios, algo que deberá a esperar al impacto de la reforma que propicia el Affordable Care Act.

Los atributos clave de la CER son:

- *Enfocada hacia la efectividad.* En las investigaciones sobre eficacia, como es el caso de los ensayos clínicos aleatorizados “explicativos” con fármacos que se realizan para obtener la aprobación de la Agencia de Alimentos y Medicamentos de EEUU (FDA) y en Europa (EMA), la pregunta más común es si el tratamiento resulta eficaz en un “entorno ideal”. En contraste, la IEC se centra en el estudio de pacientes en situaciones del “vida real” y en la toma de decisiones de la práctica clínica cotidiana. Para ello se promueven fuentes de evidencia complementarias, en especial los ensayos clínicos “pragmáticos”, los estudios observacionales –que trasladan a la realidad los resultados de los ensayos clínicos tradicionales- y las revisiones sistemáticas. Posteriormente, el proceso debería completarse con investigaciones sobre la implantación de cambios de las prácticas clínicas (“Evidence Translation”).
- *Comparación directa de alternativas efectivas.* La CER promueve las comparaciones directas entre las tecnologías sanitarias existentes -que hayan demostrado su eficacia- y las nuevas tecnologías que llaman a la puerta del mercado. Por ejemplo, la hipertensión arterial es un problema de salud con enorme relevancia social, sanitaria y económica, pero no disponemos de ensayos clínicos ni de estudios observacionales que comparen la efectividad entre los principales grupos terapéuticos disponibles y los más utilizados para el tratamiento.

- *Centrada en el paciente.* Los resultados de los estudios de eficacia no son necesariamente generalizables, tienen mucha validez interna pero poca externa. Sin embargo, lo que necesitamos conocer en la práctica es qué tratamiento es la mejor opción para un paciente determinado o un grupo de pacientes. Siguiendo con el ejemplo de la hipertensión arterial, no es lo mismo el tratamiento de un joven o un paciente mayor, un hombre o una mujer, una persona de raza afroamericana o asiática, alguien con hipertensión sin daño de órgano diana o con nefropatía. La idea es poder adaptar la decisión clínica lo más posible a las necesidades y preferencias de los pacientes. Por ello, algunos autores se refieren a la CER como “investigación de la salud centrada en el paciente” o “investigación de resultados centrados en el paciente”.

El objetivo primordial de la CER sería brindar información que ayude a elegir la opción que más se adecue a las necesidades y preferencias personales de cada paciente, y que permita priorizar las intervenciones más coste-efectivas. Médicos y pacientes necesitan saber no sólo si un tratamiento es en promedio efectivo, sino también qué intervenciones funcionan mejor para tipos específicos de pacientes (por ejemplo, los ancianos o los pacientes con múltiples enfermedades crónicas).

Planificadores y profesionales de la salud pública precisan conocer cuáles son los mejores enfoques para, por ejemplo, ofertar actividades preventivas a aquellos ciudadanos que no acceden con facilidad a la asistencia sanitaria. Este tipo de información es pues fundamental para traducir los nuevos descubrimientos en mejores resultados para la salud, acelerar la implantación de innovaciones beneficiosas y brindar el tratamiento adecuado al paciente adecuado y en el momento adecuado.

Existen múltiples superposiciones con las metodologías empleadas por la Medicina Basada en la Evidencia (MBE) y la Evaluación de Tecnologías Sanitarias. O por qué no decirlo abiertamente, quizás surja como un intento de reproducir en EE.UU. la modélica labor que el NICE (National Institute for Health and Clinical Excellence) ha desarrollado desde 1999 en Inglaterra y Gales. Tal vez por eso muchos investigadores europeos perciben a la CER como una nueva etiqueta para algo ya conocido.

El texto del Profesor Horga cubre plenamente el objetivo de analizar qué es y para qué sirve la Investigación en Efectividad Comparada. Recorre el camino de ida y vuelta que va de la eficacia a la efectividad, señalando las dificultades que se encuentran en su curso, pero también sus grandes ventajas y el enorme potencial que la CER tiene en la mejora del cuidado de los pacientes. Estoy seguro que aprovecharán su lectura.

**Dr. Juan E. del Llano Señarís**

*Director*

*Fundación Gaspar Casal*

## Principales abreviaciones (abreviaturas, siglas y acrónimos)

---

- AEMPS: Agencia Española de Medicamentos y Productos Sanitarios
- AHRQ: Agency for Healthcare Research and Quality (USA)
- ARRA: The American Recovery and Reinvestment Act
- AVAC: Años de Vida ajustados por la Calidad de Vida (Quality Adjusted Life-Year; QALY)
- BPC: Buena Práctica Clínica (Good Clinical Practice; GCP)
- CER: Investigación en Efectividad Comparada (Comparative Effectiveness Research)
- CONSORT: Consolidated Standards of Reporting Trials
- ECAP: Ensayos Clínicos Aleatorizados Pragmáticos
- EMA: Agencia Europea de Medicamentos (European Medicines Agency)
- FDA: Food and Drug Administration de Estados Unidos de América
- GTE: Guías Terapéuticas basadas en pruebas (Evidence-Based Guidelines)
- IOM: Institute of Medicine (USA)
- MBE: Medicina basada en pruebas (Evidence-Based Medicine)
- NICE: The National Institute for Health and Clinical Excellence
- NIH: National Institutes of Health (USA)
- OMS: Organización Mundial de la Salud (World Health Organization;WHO)
- PCOR: Investigación de Resultados Centrados en el Paciente ((Patient-Centered Outcomes Research)
- PCORI: Instituto de Investigación de Resultados Centrados en el Paciente (Patient-Centered Outcomes Research Institute)
- PRECIS: Pragmatic–Explanatory Continuum Indicator Summary
- QOL: Calidad de vida (Quality of life)
- RAM: Reacciones Adversas a Medicamentos
- SNS: Sistema Nacional de Salud (España)
- STROBE: Strengthening the Reporting of Observational Studies in Epidemiology
- UE: Unión Europea
- USA: Estados Unidos de América



## I. Antecedentes, conceptos y contextos

---

***“Comparative effectiveness research is the conduct and synthesis of research comparing the benefits and harms of different interventions and strategies to prevent, diagnose, treat and monitor health conditions in “real world” settings. The purpose of this research is to improve health outcomes by developing and disseminating evidence-based information to patients, clinicians, and other decision-makers, responding to their expressed needs, about **which interventions are most effective for which patients under specific circumstances**”<sup>1</sup>.***

En el año 2009 el Congreso de Estados Unidos (USA) aprobó un paquete de estímulos económicos para el país, denominado *The American Recovery and Reinvestment Act* (ARRA)<sup>2</sup>. Este paquete incluía 1.100 millones de dólares para la implementación de la investigación clínica de efectividad comparada (Comparative Effectiveness Research, en adelante CER), descrita como se ha expuesto al inicio en un informe elaborado para el Presidente y el Congreso<sup>1</sup>, y que conllevó una enmienda al Acta de la Seguridad Social de USA<sup>3</sup>. La definición indicada en ese informe, y en el ARRA, ni es novedosa ni introduce nuevos conceptos. Pero sí indica la decisión política de aportar los recursos necesarios e impulsar las medidas precisas para avanzar firmemente en esa dirección.

¿Qué aportan los estudios de CER? ¿Suponen un potencial avance en el conocimiento de la utilidad de una intervención terapéutica, o puede considerarse un paso de escaso interés? ¿Por qué es necesario o conveniente impulsar su desarrollo? ¿La información disponible o accesible hasta ahora sobre las intervenciones terapéuticas existentes es insuficiente? Para entender lo que lleva consigo esta decisión, la necesidad de inversiones económicas en su desarrollo, y la mejora que puede suponer para la calidad y eficiencia de los servicios sanitarios, hemos de ir hacia atrás en el tiempo y profundizar en la evolución de la medicina socialmente beneficiosa construida sobre bases científicamente sólidas. A lo largo de esta monografía tomaremos como elemento-guía la terapéutica farmacológica, ya que su desarrollo ha sido el motor del avance científico de la investigación terapéutica. Los principios y criterios aplicados a la terapéutica farmacológica pueden extrapolarse directamente a cualquier otra forma de intervención terapéutica.

La terapéutica farmacológica utiliza medicamentos para conseguir sus objetivos. El diccionario de la lengua española define medicamento como *“sustancia que, administrada interior o exteriormente a un organismo animal, sirve para prevenir, curar o aliviar la enfermedad y corregir o reparar las secuelas de ésta”*<sup>4</sup>. Esta definición, con escasas modificaciones, se puede leer en la mayor parte de los libros de texto de Farmacología y Medicina.

Pero estas definiciones académicas resultan poco vitales desde el punto de vista de la salud y la asistencia sanitaria. Es mucho más rica, comprometida y vigorizante la definición conceptual expuesta en Nairobi en 1985 por la Organización Mundial de la Salud (OMS)<sup>5</sup>: un medicamento *está compuesto por una sustancia química más una información pertinente*; y esta información pertinente debe

permitir en primer lugar que los medicamentos se utilicen de modo racional. Para la OMS el término *uso racional* significaba lo siguiente: que los medicamentos sean adecuados para un paciente concreto, estén disponibles en el momento que se requiere y a un precio asequible, se administren a las dosis e intervalos correctos durante el tiempo preciso, y deben ser efectivos, seguros y eficientes<sup>5</sup>. Y respecto a la *información pertinente* se indicaba que es aquella que, por una parte, permite al prescriptor elegir entre distintos tratamientos disponibles, y por otra parte fundamenta y facilita el uso racional del tratamiento que finalmente ha elegido. Para la OMS, los gobiernos son los responsables últimos de proveer esta información<sup>5</sup>. Como se puede observar, es patente la cercanía entre el concepto de medicamento de la OMS y la reciente definición de CER señalada antes, aunque las separan un cuarto de siglo.

Si la información que acompaña a una sustancia química es lo que la define y cualifica como medicamento, y permite el uso racional de éste, es fundamental entender dos aspectos esenciales: el primero es cómo se genera y obtiene una información válida disponible. El segundo es quién la analiza y elabora las conclusiones pertinentes, aunque ya hemos visto que el uso racional establece que en último término el que decide debe ser el médico prescriptor. De la suma de ambas actividades se deriva que el uso de los medicamentos sea racional, el más adecuado para ser administrado a un paciente concreto, y con la pauta terapéutica precisa. La manipulación de la información, las decisiones derivadas de análisis incorrectos por errores metodológicos o conceptuales, o por intereses espurios (sean comerciales, económicos o profesionales), pueden generar un terrible impacto negativo en la salud de la población. Evaluaré ambos apartados, siguiendo el proceso evolutivo del desarrollo de la investigación sobre los efectos de los medicamentos en los tiempos recientes.

**Como método de generación de información, sólo son válidos los ensayos clínicos: El momento y el lugar de la decisión**

En 1956 La compañía farmacéutica alemana Grünenthal comercializó la talidomida como un nuevo sedante, presentando como argumento a su favor que era más seguro que el fenobarbital. La compañía farmacéutica señalaba que, además, era útil para controlar las náuseas matutinas de las embarazadas. La talidomida se comercializó en más de 40 países, y en 1960 la licenciataria W. S. Merrel Company presentó ante la Food and Drug Administration (FDA) la solicitud de autorización para comercializar Kevadon (talidomida) en USA. La funcionaria médico contratada para evaluar el dossier, F. Kelsey, rehusó aprobar la solicitud, debido a la escasez de pruebas de eficacia y de seguridad en tratamientos a largo plazo del producto. Mientras tanto la compañía distribuyó en USA el producto a unos 20.000 pacientes, incluidas mujeres embarazadas. A finales de 1961 se demostró fuera de USA la asociación entre el uso de talidomida y diversas malformaciones congénitas, entre ellas miles de casos de una malformación rara, la focomelia (deficiencia en el desarrollo osteomuscular de las extremidades). La FDA identificó en USA 17 casos de malformaciones congénitas asociadas al uso de talidomida<sup>6</sup>.

El impacto de esta información, difundida por la prensa y la televisión, fue tremendo en la sociedad norteamericana. La amenaza de que se pudiesen comercializar medicamentos cuyo beneficio terapéutico no fuese adecuado, que presentasen un riesgo elevado de producir efectos adversos graves e irreversibles, e incluso que ese hecho estuviese sucediendo ya, dio lugar a la aprobación por el Congreso

de USA en 1962 de unas propuestas cuyo refrendo legal había sido discutido y postergado durante años, las denominadas *enmiendas Kefauver-Harris*. Según esta normativa, antes de comercializar un nuevo medicamento, el laboratorio titular debía *demostrar la eficacia* del medicamento, y notificar cualquier efecto adverso grave tras su comercialización<sup>7</sup>.

Un elemento clave en esta enmienda era el requisito condicionante que se establecía como necesario para demostrar la eficacia de un nuevo medicamento: las pruebas de eficacia debían basarse en adecuados ensayos clínicos controlados, realizados por profesionales expertos cualificados<sup>7</sup>. Además, los sujetos del estudio debían haber prestado previamente su consentimiento informado para el mismo. A partir de este momento se fijaba el principio de que sólo los resultados de los ensayos clínicos controlados pueden considerarse pruebas adecuadas para sustentar la eficacia del medicamento, y todos los demás tipos de estudios (estudios epidemiológicos observacionales, estudios basados en la experiencia y seguimiento de casos, etc.) no eran aceptables con esta finalidad. Este precepto se vio reforzado cuando se llevó a cabo un estudio, promovido por la nueva normativa, que revisó las pruebas de eficacia disponibles para todos los medicamentos que habían sido autorizados previamente por la FDA entre 1938 y 1962, aprobaciones cuya base solía sustentarse sobre todo en informes de seguridad, datos de expertos y distintos tipos de estudios clínicos. Se evaluaron cerca de 4.000 formulaciones, con 16.000 indicaciones autorizadas. La revisión concluyó que sólo el 12% de los medicamentos podían considerarse eficaces en todas sus indicaciones, y en el 40% de las indicaciones se consideró que los medicamentos eran “menos que eficaces”. Tras las alegaciones y envío de nuevos informes y estudios por las empresas titulares, se concluyó finalmente que un tercio de los medicamentos autorizados entre 1938 y 1962 eran ineficaces<sup>8</sup>. Otro aspecto muy relevante de las conclusiones, y no menos importante que el anterior, fue la percepción de que los datos de estudios con limitada validez interna y/o externa, los conflictos de interés y los entresijos administrativos constituían los obstáculos más relevantes en el proceso de evaluación de los informes sobre la eficacia de los medicamentos<sup>9</sup>.

Como se ha señalado antes, la normativa establecía que a partir de ese momento las pruebas de eficacia de los nuevos medicamentos debían basarse en los resultados de adecuados ensayos clínicos controlados. ¿Qué se entiende por ensayo clínico? La Directiva 2001/20/CE del Parlamento Europeo define para la Unión Europea (UE) los ensayos clínicos como: “Toda investigación efectuada en seres humanos dirigida a determinar o verificar los efectos clínicos, farmacológicos y/o demás efectos farmacodinámicos, y/o identificar cualquier reacción adversa, y/o estudiar la absorción, distribución, metabolismo y excreción de uno o varios medicamentos en investigación con el propósito de determinar su seguridad y/o eficacia”<sup>10</sup>. La normativa española vigente transcribe literalmente esta definición<sup>11</sup>. El reciente Reglamento del Parlamento Europeo de 2014 aplica esta definición a los “estudios clínicos”, matizando posteriormente el concepto de ensayo clínico a una parte de los estudios, que serán siempre ensayos clínicos cuando se trate de medicamentos no comercializados<sup>12</sup>.

Ya en el título de la Directiva 2001/20/CE se indicaba que estas normativas y directivas comunitarias tienen como finalidad trasladar a disposiciones legales de seguimiento obligatorio una serie de acuerdos que ya habían sido consensuados y refrendados mediante una firma conjunta en 1995 en la conferencia

internacional de armonización entre USA, la UE y Japón<sup>13</sup>. Estos acuerdos son resultado de la aceptación de algunos principios de las enmiendas aprobadas en USA para la FDA, y establecen, desde 1995 y de modo obligatorio para todas las partes firmantes, que, en estas tres regiones del mundo, el desarrollo de nuevos medicamentos ha de seguir un proceso basado en la información generada por ensayos clínicos, y que, para considerar válida esa información, y aceptable para todos ellos, la realización de ensayos clínicos ha de seguir unos procedimientos estandarizados. Estos procedimientos se denominan “Buena Práctica Clínica” (BPC)<sup>13</sup>. Las BPC se definen como un conjunto de requisitos éticos y científicos de calidad reconocidos a escala internacional, que *deben cumplirse en la planificación, la realización, el registro y la comunicación de los ensayos clínicos en que participen seres humanos*<sup>13</sup>. El cumplimiento de esas normas, además de garantizar la seguridad y el bienestar de los sujetos participantes, *garantiza la validez y credibilidad de los datos del ensayo clínico*. Se facilita de este modo la aceptación mutua de datos clínicos por parte de las autoridades reguladoras de estas jurisdicciones, y “*se deberá seguir esta guía cuando se generen datos de ensayos clínicos que se pretendan presentar a las autoridades reguladoras*” para la autorización de un nuevo medicamento<sup>13</sup>.

Así pues, a partir de ese momento el proceso queda fijado y los criterios cerrados. Para que se autorice un nuevo medicamento sólo pueden considerarse válidos y aceptables para las autoridades reguladoras, y para la mayor parte de científicos y evaluadores, los datos generados por la metodología de investigación que siga los requisitos científicos (y éticos) establecidos para los ensayos clínicos. Sólo los datos de los ensayos clínicos controlados son fiables y admisibles, y la información sobre los nuevos medicamentos o nuevas indicaciones terapéuticas ha de basarse necesariamente en ellos.

### **La justificación de la decisión: Las virtudes de los ensayos clínicos controlados y aleatorizados**

¿Por qué se considera que las únicas pruebas científicamente válidas para establecer si un medicamento es eficaz y seguro deben provenir de ensayos clínicos controlados? ¿En qué se diferencian metodológicamente los ensayos clínicos y el resto de estudios que se realizan en humanos? La contestación a esta pregunta es fundamental en la aproximación a la CER y a cualquier tipo de soporte de las decisiones clínicas, y debe ser previamente y cuidadosamente considerada. En caso contrario, será difícil entender la situación, las limitaciones y los conflictos en que nos encontramos y movemos actualmente y por qué y para qué se precisan ciertos cambios profundos.

La estructura de un ensayo clínico debe ser tal que permita, sin dudas razonables de tipo científico, llegar a la siguiente conclusión: la administración de un medicamento produce un efecto definido e identificado, cuya magnitud o intensidad puede expresarse de modo cuantitativo. En consecuencia, puede decirse que ese medicamento posee una eficacia objetiva. Para que esta conclusión sea posible, el experimento (ensayo clínico) debe estar diseñado de tal modo que la medida de los efectos (respuesta) sea válida y validable, y que haya una *sóla y única variable* de intervención en el estudio a la que atribuir las modificaciones detectadas en la medida de la respuesta; esta única variable explicativa debe ser la intervención terapéutica, el medicamento administrado. Sólo en estas condiciones podemos concluir que ese medicamento ejerce un efecto determinado, y estaría justificada la autorización de su comercialización con la finalidad de obtener ese efecto.

¿Cómo se debe diseñar, por lo tanto, un ensayo clínico, de tal modo que permita establecer la conclusión científica pertinente en el sentido antes expuesto?

### **Objetivo primario**

En primer lugar es necesario definir con precisión cuál es el efecto que se elige para cuantificar la magnitud de la respuesta al medicamento. El denominado objetivo primario (por lo general denominado *primary end point* en inglés), es el elemento clave de cualquier ensayo clínico, como de cualquier estudio de investigación. Define exactamente cuál es la pregunta a la que se pretende contestar con el estudio; y por lo tanto establece la respuesta que se va a medir en los pacientes durante el estudio como variable que identificará la eficacia de un medicamento o de una intervención terapéutica, cuantificando sus efectos en ese tipo de pacientes. Es el elemento que condiciona el diseño metodológico del estudio<sup>14-16</sup>.

El objetivo primario no solo constituye el elemento básico de un ensayo clínico. Desde el punto de vista de la asistencia sanitaria, el objetivo primario identifica con precisión cuál es el beneficio que podré obtener en los pacientes a los que administre el medicamento, como consecuencia de su administración. En este sentido, en consecuencia, los objetivos primarios deberían idealmente ser variables de evolución clínica de relevancia primaria para los pacientes (supervivencia, calidad de vida), o de gran relevancia (incidencia de infarto agudo de miocardio, riesgo de ictus, necesidad de hospitalización o de intervenciones quirúrgicas...). Pero muy frecuentemente los objetivos primarios son medidas subrogadas o vicariantes: biomarcadores u otras medidas de laboratorio (colesterol, glucemia), medidas de exploraciones clínicas (presión arterial, volumen espiratorio forzado), o bien variables clínicas de respuesta alternativas a las verdaderamente importantes, cuya importancia desde el punto de vista de la relevancia para la salud de los pacientes podría considerarse secundaria o limitada<sup>16,17</sup>. En muchas ocasiones, el análisis de los objetivos primarios de distintos ensayos clínicos de un medicamento o medicamentos similares, relacionados con una misma situación o patología concreta nos muestra que dichos objetivos son heterogéneos, diferentes para distintos ensayos en la misma población, y podrían catalogarse como carentes de verdadera relevancia y consistencia, y sanitariamente opacos<sup>18</sup>.

Como consecuencia de su carácter esencial, el objetivo primario es también el que se utiliza como elemento determinante en el cálculo del tamaño muestral. Un ensayo clínico debe incluir el número de sujetos necesario para poder detectar diferencias entre dos tratamientos estadísticamente significativas, si es que dichas diferencias existen. Y por lo tanto limitar (por lo general hasta una probabilidad menor del 20%) el denominado error beta, consistente en que como conclusión del estudio se establezca que no hay diferencias entre el tratamiento experimental y el de referencia, cuando en realidad esas diferencias existen.

El objetivo primario va a ser, como veremos más adelante en detalle, el apartado del estudio que más frecuentemente se intente manipular y falsear, sobre todo cuando se trate de extrapolar la eficacia de un medicamento (por ejemplo, su eficacia para reducir el colesterol plasmático o la glucemia plasmática) a pretendidos y no demostrados efectos beneficiosos sobre la salud (en los casos citados: reducción de morbilidad cardiovascular, en base a pretendidas relaciones (no) existentes entre ambos procesos). En los ensayos clínicos en patologías cardiovasculares, muchos autores opinan que los objetivos primarios que no sean la mortalidad pueden inducir a error a los profesionales y responsables de gestión

asistencial sobre la eficacia global de la intervención de que se trate<sup>19</sup>. Esto sucede en particular cuando se han elegido para los ensayos clínicos objetivos primarios subrogados o solo de relativa importancia para la salud de los pacientes.

### **Objetivos secundarios**

Los objetivos secundarios están constituidos por otros elementos de respuesta del estudio que pueden recogerse durante el desarrollo de éste, pero que no han sido los condicionantes del diseño del estudio; el ensayo no se ha diseñado para responder a éstos objetivos. La confusión entre la validez de los objetivos primarios y los secundarios, por ignorancia, hábitos erróneos de trabajo, o pretendida deliberadamente por interés comercial o personal, también se constituye frecuentemente en elemento susceptible de manipulación y distorsión. Un ensayo clínico no está diseñado para responder a los objetivos secundarios con un nivel de fiabilidad adecuado, si bien los datos de éstos pueden analizarse para generar nuevas hipótesis o valoraciones críticas de los resultados del objetivo primario, así como para reforzar la credibilidad de estos resultados o establecer su plausibilidad biológica o patológica<sup>14,17</sup>. Por otra parte, la falta de diferencias estadísticamente significativas entre dos tratamientos respecto a un objetivo secundario simplemente puede derivar de que el tamaño muestral es insuficiente para detectar dichas diferencias.

Típicamente, los acontecimientos adversos recogidos durante el ensayo clínico, o la tolerabilidad al medicamento en estudio, debido a la dificultad para establecerlos como objetivo primario, son objetivos secundarios del mismo, así como aspectos de beneficio terapéutico de importancia secundaria. Pero no es infrecuente que los objetivos secundarios incluyan los de mayor interés para el paciente, el médico y el sistema de salud, y que deberían haber sido los objetivos primarios del estudio: morbilidad de alta relevancia, calidad de vida o incluso mortalidad. Pondré algún ejemplo en este sentido para dimensionar su trascendencia a la hora de elaborar recomendaciones terapéuticas.

**Caso 1:** Levosimendan es un medicamento inotrópico que aumenta la afinidad de la troponina C por el calcio libre intracelular con características únicas desde el punto de vista farmacodinámico y farmacocinético. Esta propiedad podría incrementar el efecto contráctil del calcio sin aumentar el consumo de  $O_2$ , y ser por lo tanto muy útil en el tratamiento de la insuficiencia cardíaca<sup>20</sup>. Uno de los estudios principales (los llamados “pivotales”, aquellos que aportan los datos fundamentales de eficacia y seguridad del medicamento) en los que se basó su comercialización, denominado LIDO<sup>21</sup>, demostró que levosimendan alcanzaba su *objetivo primario* (mejoría hemodinámica clínicamente relevante) de modo favorable y estadísticamente significativo respecto al tratamiento con dobutamina. Como *objetivo secundario* del estudio se cuantificó la mortalidad de los pacientes a 180 días: en el grupo tratado con levosimendan falleció el 28% de pacientes, frente al 36% en el grupo de dobutamina<sup>21</sup>. Con posterioridad, otros estudios, como el denominado SURVIVE<sup>22</sup>, han evaluado la *mortalidad como objetivo primario* en pacientes tratados con levosimendan frente al tratamiento con dobutamina, sin detectar diferencias entre uno y otro tratamiento<sup>22</sup>. No es infrecuente que los datos de una variable evaluada como objetivo secundario en un estudio, sean claramente distintos cuando se diseña el estudio para su evaluación específica como objetivo primario, ya que en estas circunstancias se definen mejor las variables, se establecen métodos adecuados para recoger

los eventos en todos los pacientes del estudio, se tienen en cuenta las pérdidas de pacientes, etc.

**Caso 2:** En 1998 se publicó un ensayo clínico (the ABCD trial) cuyo objetivo primario era valorar, en pacientes con hipertensión arterial y diabetes mellitus tipo 2, si el tratamiento con nisoldipino o enalapril podía mejorar la evolución del aclaramiento de creatinina en estos pacientes<sup>23</sup>. Aunque ambos tratamientos conseguían una respuesta similar en las variables secundarias de control evolutivo (presión arterial, hemoglobina glucosilada, colesterol plasmático, etc), el riesgo de padecer un infarto fue 7 veces mayor en el grupo tratado con nisoldipino que en el grupo que recibió enalapril<sup>23</sup>. Sin embargo, los autores no aceptan la validez de este hallazgo. En el artículo, y lo resaltan en el abstract, los autores argumentan que el estudio se diseñó para evaluar como objetivo primario si la reducción intensiva de la presión arterial, con uno u otro medicamento, generaba disminución de los eventos asociados a la diabetes, respecto a la reducción no intensiva de la presión arterial. Y que, como los hallazgos sobre el riesgo incrementado de infarto se basan en un objetivo secundario del estudio, no pueden aceptarse sin confirmación<sup>23</sup>.

Estos dos casos nos indican cómo se descarta con facilidad por los promotores y los investigadores la fiabilidad de los hallazgos relacionados con los objetivos secundarios de los ensayos clínicos cuando ésta decisión les interesa, y que realmente pueden tener razón: extraer conclusiones sobre los objetivos secundarios puede ser una equivocación manifiesta, como se demuestra cuando se tratan de confirmar los resultados de un objetivo secundario en otro ensayo donde pasa a ser objetivo primario, como en el caso de levosimendan, a pesar de tratarse de un objetivo secundario constituido por una variable “dura” (de difícil error de interpretación) como la muerte. Después veremos que, a la hora de establecer recomendaciones terapéuticas, muchos autores de meta-análisis o revisiones sistemáticas dejan aparte estos criterios sin ningún pudor, cuando conviene a sus intereses.

### **Sesgos y errores de interpretación**

El diseño del estudio, a partir del momento en que se elige el objetivo primario, debe ser tal que permita extraer conclusiones válidas para ese objetivo; es decir, el diseño debe dar validez (validez interna) a los datos generados en el estudio *en relación con ese objetivo primario*. Un estudio estará bien diseñado, por lo tanto, si permite extraer conclusiones válidas en relación con el objetivo primario del mismo. Un buen diseño, en consecuencia, tratará de controlar rigurosamente aquellas situaciones o circunstancias que puedan hacer que los datos finales estimados para los parámetros de respuesta elegidos sean diferentes a los valores reales, o que las conclusiones extraídas sean erróneas. Es decir, al planificar el ensayo clínico se deben controlar los sesgos potenciales del experimento, los errores de diseño que puedan dar lugar a resultados o interpretaciones erróneas.

### **El ensayo controlado**

El error más preocupante en un estudio en el que se pretende establecer la eficacia de un medicamento es el de una falsa *atribución de causalidad*: establecer una relación de causalidad errónea entre la intervención (tratamiento) y la respuesta. Un ensayo clínico controlado es el modo científico más riguroso que existe para establecer una sólida relación de causalidad entre el medicamento y la *respuesta al objetivo primario*. Para ello, el ensayo clínico debe descartar otras variables

(edad, evolución de la enfermedad, patologías o tratamientos concomitantes, cambios en los hábitos alimenticios, laborales o de salud, del paciente, cambios en la actitud del paciente hacia su enfermedad, etc.), que pudiesen ser la causa, total o parcial, del cambio detectado en las variables medidas. Esta implicación de otras variables es prácticamente imposible de cuantificar (y en ocasiones es desconocida), por lo que el ensayo clínico controlado utiliza una metodología que elimina la participación de otras variables en el resultado final: asigna a los pacientes del estudio a dos tratamientos distintos: el que se quiere evaluar en el ensayo, y el tratamiento estándar hasta ese momento para la población en estudio. Si los dos grupos son similares, la única diferencia entre los dos grupos, el experimental y el control, es el tratamiento que se evalúa en el estudio, y el valor de la medida de la variable primaria es diferente en ambos grupos al finalizar el estudio, la única justificación de que exista una diferencia en el resultado es la diferencia en la intervención terapéutica recibida por los distintos grupos. La conclusión, por lo tanto, es que el tratamiento puesto a prueba es diferente al tratamiento control. Si no existe un tratamiento estándar probado previamente, o si se requiere por razones metodológicas y científicas para determinar la eficacia y la seguridad de una intervención, el tratamiento control puede ser una sustancia farmacológicamente inerte e indiferenciable del tratamiento activo, que se denomina placebo, o una simulación de la intervención (por ejemplo, en cirugía o en acupuntura)<sup>24-26</sup>.

No son científicamente aceptables los resultados de estudios no controlados, y éstos deben ser sólo considerados útiles con finalidades descriptivas de la experiencia o generadoras de hipótesis, pero nunca válidos con el fin de establecer una relación de causalidad. Por lo tanto es totalmente carente de base científica esperar que en los estudios controlados se obtendrán los mismos resultados que se han obtenido en estudios no controlados, cuando se evalúe el nuevo tratamiento en nuevos pacientes.

### Aleatorización

Sin embargo, ¿cómo puedo garantizar que el grupo tratado con el medicamento experimental y el grupo control *solamente* se diferencien en el tratamiento recibido, y que ambos grupos sean similares en el resto de posibles factores influyentes en la respuesta? Si no se puede controlar este sesgo, es probable que se obtenga una conclusión equivocada sobre la atribución de causalidad, y por lo tanto el resultado del estudio no sería aceptable. El error sistemático por el que los pacientes asignados al grupo experimental y al grupo control difieren entre sí en condiciones basales (antes del inicio del tratamiento) se denomina sesgo de selección (por los metodólogos de ensayos clínicos), o sesgo de asignación (por los estadísticos).

La aleatorización o distribución al azar es el procedimiento que permite de modo más fiable que las variables que afectan a la respuesta se distribuyan de modo similar en los dos grupos de estudio, reduciendo el riesgo de falta de homogeneidad en condiciones basales y haciendo comparables a los dos grupos en todo excepto en la variable de exposición (la intervención terapéutica). Si estas diferencias se minimizan, las diferencias obtenidas en la respuesta pueden aceptarse como resultado de la diferente intervención terapéutica, no de las diferentes características basales de los pacientes. Y, por lo tanto, como se comentaba antes, sí que puede establecerse una relación causal entre el tratamiento administrado

y la diferente respuesta obtenida. Los resultados del estudio obtenidos poseen entonces una elevada validez interna<sup>27-29</sup>.

Ningún otro tipo de estudio clínico diferente a los ensayos clínicos puede seguir un proceso de aleatorización, ya que en el resto de estudios (estudios observacionales), los pacientes son incluidos en los estudios por el hecho de que su médico ha decidido, previamente a su inclusión en el estudio, administrarles (o no), un tratamiento concreto, porque el médico había juzgado previamente que ese tratamiento era el más adecuado para ese paciente concreto en su situación concreta; en el ensayo clínico, por el contrario, los sujetos del estudio reciben un tratamiento por el hecho de participar en el estudio, y la asignación a uno u otro tratamiento puede aleatorizarse. Por lo tanto, ningún otro tipo de estudio puede garantizar la homogeneidad de los grupos en los que se compara el efecto de la intervención como se hace en un ensayo clínico, y en consecuencia sólo la validez interna del ensayo clínico es óptima, y netamente superior a otros diseños de estudios. Por ello las administraciones sanitarias han establecido que la información sobre los nuevos medicamentos o nuevas indicaciones terapéuticas ha de basarse necesariamente en ensayos clínicos.

Existen otras garantías de fiabilidad que pueden incorporarse, y por lo general así lo hacen, en el diseño de los ensayos clínicos, aumentando la validez interna de los resultados obtenidos. No nos extenderemos más en este terreno, aunque es necesario, por su importancia, hacer una referencia al enmascaramiento del tratamiento, en formato simple ciego o doble ciego, que permite controlar una sustancial fuente de error<sup>14,15,30</sup>, la relacionada con el juicio clínico y la recogida de información no sesgada, casi insoslayable en la recogida de datos de estudios observacionales.

No dudamos, por lo tanto, en afirmar que los ensayos clínicos doble ciego, controlados y aleatorizados, cuando se realizan y analizan correctamente, aportan de modo incuestionable la mayor seguridad posible sobre la validez interna de los resultados obtenidos en una investigación clínica, en lo que se refiere al beneficio relacionado con el objetivo primario de la intervención. Y más aún si se confirma con los resultados de otros estudios similares. Así pues, son el *gold standard*, el patrón de referencia utilizado para demostrar o refutar el beneficio de una intervención terapéutica.



## II. Aproximaciones a la efectividad comparada sin datos de estudios de efectividad comparada. Aportaciones y limitaciones

---

Una vez que las empresas titulares de un medicamento disponen de lo que ellas consideran pruebas suficientes del beneficio terapéutico de un medicamento, envían a las agencias reguladoras los datos de eficacia obtenidos en los ensayos clínicos. Las agencias evalúan los informes de los ensayos y, si los consideran adecuados, autorizan la comercialización del medicamento evaluado. Tal y como establecían las enmiendas Kefauver-Harris y los acuerdos internacionales de consenso recogidos en las normas BPC, para que un medicamento sea comercializado, es necesario que las pruebas de su eficacia se basen en adecuados ensayos clínicos controlados, realizados por profesionales expertos cualificados<sup>7,13</sup>, y aceptamos tales pruebas como válidas. *Los ensayos clínicos han cumplido con su finalidad.*

Además de los informes que las agencias reguladoras emiten tras la evaluación de los datos remitidos por las empresas, informes a los que se puede acceder libremente, estos estudios se publican en revistas científicas tras su revisión por expertos, lo que confiere aún una mayor garantía de la adecuación metodológica y la validez del experimento realizado. Como todo experimento científico, en su publicación se aporta la información necesaria y pertinente sobre el experimento realizado y los resultados obtenidos, con el fin de que puedan ser discutidos, reproducidos, confirmados o rechazados. Este es el modo de generar y avanzar en el conocimiento científico, empleando los procedimientos de mayor nivel de transparencia, exigencia y garantía para la comunidad científica internacional. Los avances en el conocimiento relacionado con la salud así lo exigen.

Pero, si bien los ensayos clínicos son poco criticables en relación con la finalidad para la que han sido diseñados (aportar pruebas con elevada validez interna de la eficacia de un medicamento), sí presentan limitaciones respecto al proceso inmediatamente siguiente, y que avanzábamos al principio. La información que poseemos del nuevo medicamento es la **necesaria** para su autorización, pero ¿es **suficiente y pertinente** la información ahora disponible? Con los datos disponibles, ¿puede el prescriptor elegir entre los distintos tratamientos disponibles, y puede fundamentar el uso racional del tratamiento que finalmente ha elegido?

### **Resultado del ensayo: eficacia comparada vs respuesta al tratamiento**

Antes de continuar, es necesario dejar claro otro concepto respecto al análisis de los resultados de los ensayos clínicos: los ensayos clínicos controlados permiten, una vez recogidas las medidas de la variable relacionada con el objetivo primario, comparar las respuestas obtenidas en los grupos del estudio. Este procedimiento permite concluir si las *diferencias* encontradas son o no estadísticamente significativas, y el posible error de extrapolación de esas diferencias (error alfa o tipo I) a todo el universo del estudio; pero no pueden establecerse conclusiones sobre el *valor absoluto* de la respuesta medida en el estudio. De hecho, los valo-

res absolutos varían de un estudio a otro, ya que, como se ha indicado antes, en dicha respuesta participan también otros elementos no relacionados con el medicamento en estudio: edad, sexo, tiempo de evolución de la enfermedad, hábitos higiénico-dietéticos, cambios en el estilo de vida, otros tratamientos concurrentes, calidad de la asistencia sanitaria prestada... ¿Cómo puede saberse cuál es el valor absoluto del efecto de un medicamento en este contexto? No se puede. Sólo puede establecerse el valor de la diferencia entre los dos grupos. Como se ha expuesto antes, si la única diferencia entre los grupos es el tratamiento, la diferencia en la respuesta puede atribuirse a la diferencia en el tratamiento. Erróneamente, y pue que intencionadamente en ocasiones, a veces se valora el efecto absoluto de la respuesta a un tratamiento para “estimar” el impacto de dicho tratamiento sobre la salud de los pacientes a los que se administre este medicamento en la práctica asistencial, sin tener en cuenta el resto de factores que han podido influir en el valor absoluto de la respuesta obtenida.

La diferencia entre los resultados de una intervención terapéutica en los ensayos clínicos aleatorizados y los resultado de su utilización en el *mundo clínico real* (*real world settings* en la definición de CER<sup>1</sup>) ya fue señalada e identificada por A. Cochrane hace 40 años. Para la aplicación de una intervención clínica Cochrane introdujo una jerarquía de pruebas requeridas: Primero: ¿El tratamiento puede ser útil? Es decir, ¿ha demostrado su eficacia en circunstancias experimentales controladas? Segundo: ¿Es útil en el mundo clínico real (efectividad), más allá de las pruebas obtenidas en condiciones experimentales controladas? Tercero: ¿Cuál es su valor para alcanzar esa efectividad (eficiencia)? El problema que nos ocupa en esta monografía tiene relación con la adecuada comprensión de las diferencias existentes entre los dos primeros conceptos<sup>31,32</sup>. El tercero, la eficiencia, excede el propósito de este trabajo, pero es necesario recalcar, y lo haré en alguna ocasión más, que sin tener datos adecuados de la efectividad de una intervención, por lo general el análisis de su eficiencia (análisis farmacoeconómico) se realiza incorrectamente basándose en los datos de su eficacia, directos o modelizados con mayor o menor complejidad y profundidad. Este análisis, sin datos reales de efectividad, puede ser acertado o puede ser profunda y absolutamente erróneo, y justificar decisiones que pueden ser acertadas o por el contrario totalmente equivocadas, sin que se puedan establecer criterios para distinguir el riesgo de errores en el análisis. No entender la jerarquía y la relación jerarquizada entre estos conocimientos puede llevar a tomar (¿frecuentemente?) decisiones de coste-oportunidad arriesgadas o desacertadas<sup>33</sup>.

En el contexto que nos ocupa, se define *eficacia* como el grado en que una intervención produce un efecto beneficioso en condiciones experimentales controladas, como es el marco de un ensayo clínico. Entendiendo el término grado como la diferencia en ese efecto entre el grupo experimental y el control. *Efectividad* es el grado en que una intervención produce un efecto beneficioso en las condiciones reales de la práctica asistencial habitual<sup>33,34</sup>. Cuando se realizan ensayos clínicos de efectividad comparada, el término grado hace también referencia a la diferencia en la efectividad entre las dos intervenciones comparadas.

Es un grave error asumir que los resultados de los estudios de eficacia pueden aplicarse fielmente a la práctica asistencial. Incluso si una intervención ha demostrado su beneficio en un ensayo, puede no ser adecuada para su uso clínico generalizado. La efectividad de un tratamiento está determinada por su eficacia, pero también depende de las condiciones de uso, de las características demo-

gráficas y clínicas de los pacientes a los que se administra, del cumplimiento terapéutico, de la disponibilidad del producto, o de la calidad y cobertura de los servicios asistenciales<sup>33</sup>.

En este nuevo paso que es necesario dar, para tomar decisiones sobre la efectividad y el uso racional de medicamentos en la práctica asistencial, la información disponible en las publicaciones de los ensayos clínicos presenta varias e importantes limitaciones. Destacaré, por su importancia para esta monografía, dos de ellas: las limitaciones en la extrapolación de los datos del estudio desde los pacientes del estudio a los pacientes que se tratan en la práctica asistencial, y las limitaciones en la extrapolación de los objetivos de eficacia a la práctica asistencial.

**Población en estudio.  
Validez externa de  
los datos del estudio.  
Extrapolación y  
generalización**

El análisis de estas limitaciones tiene como punto de reflexión el análisis de la validez externa de los resultados del estudio, que puede resumirse como la contestación a la pregunta: *¿a quién puede aplicarse los resultados del estudio?* Se trata de un análisis fundamental, pero complejo, que puede conllevar el uso no justificado de tratamientos en algunos pacientes, y, por lo general, éste es un punto muy descuidado en la elaboración de las recomendaciones terapéuticas, y en la reflexión sobre la potencial efectividad de una intervención terapéutica concreta<sup>35,36</sup>. Las pruebas de eficacia de un tratamiento son sólo el primer paso para establecer si dicho medicamento (o intervención preventiva o terapéutica) es adecuado para su empleo clínico habitual. Así pues, uno de los elementos críticos en este proceso es cómo extrapolar los resultados del estudio desde la muestra accesible, incluida en el estudio, a la población general.

La efectividad de un tratamiento está determinada por su eficacia, pero también está condicionada por la población realmente tratada. Los ensayos clínicos controlados *seleccionan* los pacientes en los que van a poner a prueba la eficacia del medicamento. Esta selección se realiza utilizando normas muy estrictas, denominadas criterios de inclusión y exclusión. Los principios de análisis científico establecen, con total firmeza, que la extrapolación de los resultados del estudio sólo se puede realizar de modo válido si la población diana de la práctica asistencial, alejada espacial o temporalmente respecto a la del estudio, cumple los mismos criterios de inclusión y exclusión que ésta<sup>37,38</sup>.

Los criterios de inclusión definen las características que *deben cumplir todos* los pacientes para ser incluidos en el estudio: algunos son criterios demográficos, que limitan la edad de los participantes (mayores de 18 años, de 18 a 65 años...), o el sexo. Son importantes los criterios de inclusión relacionados con las características de la patología en estudio: procedimientos diagnósticos, pruebas diagnósticas y criterios diagnósticos utilizados; estadio de la enfermedad y su gravedad; tiempo transcurrido desde el inicio del cuadro patológico o tiempo desde el diagnóstico; duración de la enfermedad; reversibilidad del proceso; respuesta a tratamientos previos (o falta de respuesta, o pérdida de respuesta, o intolerancia). Es inconcebible cómo se pasan por alto estos aspectos cuando se extrapolan los resultados del estudio a cualquier paciente solamente porque en su historia clínica figura un diagnóstico cuya denominación o código de clasificación coincide con el de los pacientes del ensayo. Es inconcebible en términos racionales y científicos, pero es lo habitual.

La población en estudio (el llamado universo del estudio), no sólo queda delimitada por los criterios anteriores. No todos los pacientes que cumplen con los

criterios de inclusión pueden ser reclutados. Existen otros criterios (criterios de exclusión), que pueden impedirles participar, porque el estudio puede presentar un riesgo para esos pacientes en particular, o porque podrían generar confusión al analizar los resultados o al valorar acontecimientos adversos. Algunos de estos criterios de exclusión habituales son la presencia otras patologías concomitantes, por lo general las de carácter crónico y sistémico, moderadas o graves. También se excluyen por lo general a los pacientes en los que se detecten alteraciones en las pruebas funcionales hepáticas, metabólicas, cardiovasculares o renales, salvo las relacionadas con la patología del estudio. Suelen ser excluidos además los pacientes tratados con medicamentos que puedan interferir con el medicamento en estudio por riesgo de interacciones farmacocinéticas o farmacodinámicas, o porque esos medicamentos puedan interferir con la medida de las variables en estudio, o generar acontecimientos adversos de difícil atribución de causalidad diferencial con respecto al medicamento en estudio. Y, por supuesto, el embarazo o la posibilidad de embarazo (salvo la garantía firmada de protegerse frente a este evento por parte del sujeto del ensayo), excepto en los escasos ensayos clínicos dirigidos a esta población.

En consecuencia, y debido a que los ensayos clínicos controlados y aleatorizados se realizan en poblaciones muy seleccionadas, la generalización de los resultados de los ensayos clínicos es uno de los pasos más sensibles en el proceso del análisis del grado de efectividad de una intervención terapéutica en relación con la delimitación de la población que puede beneficiarse de dicha intervención, sobre todo si ésta no es la misma población seleccionada en los estudios que han proporcionado las pruebas de eficacia<sup>35,36,39,40</sup>. La validez externa es frecuentemente un análisis que, si no se realiza siguiendo criterios previamente establecidos, depende mucho de la subjetividad y la perspectiva particular del individuo que realiza el examen<sup>36,40</sup>: para un médico concreto, ante un paciente concreto, pueden ser perfectamente aplicables los resultados de un ensayo clínico, mientras que para otro médico, en la misma situación, la validez externa puede ser ínfima. No entraremos aquí en qué procedimientos deberían utilizarse para realizar una aproximación racional a este problema, sino resaltar que por lo general no se protocolizan dichos procedimientos ni se utilizan criterios contrastados a la hora de elaborar recomendaciones terapéuticas generales o tomar decisiones terapéuticas en pacientes concretos. Es decir, la extrapolación de los resultados del estudio presenta alto riesgo de error.

### ***Análisis de subgrupos***

La valoración de la efectividad de las intervenciones terapéuticas en grupos particulares o sensibles de pacientes, es un elemento crítico en los objetivos del proyecto CER, por lo que es importante profundizar en este apartado. Frecuentemente, la valoración de la efectividad de un tratamiento concreto en un subgrupo de pacientes definido proviene de dos juicios erróneos: el primero, de extrapolar los datos de eficacia de un ensayo clínico a la estimación de la efectividad. El segundo, de la evaluación errónea de los datos de eficacia en ese subgrupo de pacientes.

Existen muchas limitaciones, además de las ya indicadas, para llevar a cabo correctamente la extrapolación de los resultados del ensayo clínico a la práctica asistencial diaria: la duración limitada del tratamiento en el ensayo clínico, inferior a la duración en la práctica habitual en el caso de patologías subagudas o crónicas; la falta de datos acerca de cómo podría variar la respuesta al tratamiento si estu-

viesen presentes otras enfermedades o tratamientos concomitantes, etc<sup>41</sup>. Pero dentro del apartado de la población en estudio, otro punto sensible y relevante que puede dar lugar, y lo hace frecuentemente, a errores de extrapolación, está asociado a la realización de valoraciones de los resultados del estudio no en el conjunto de pacientes evaluado (en el que se ha diseñado el estudio), sino sólo en un subgrupo de esos sujetos, que pueden ser atractivos desde el punto de vista práctico por las características diferenciales que presentan. Los estadísticos y epidemiólogos alertan sobre los graves errores que se cometen en los análisis de subgrupos, pero los clínicos señalan que también puede ser erróneo aplicar el resultado del estudio al conjunto de la muestra evaluada, sin tener en consideración las diferencias entre subgrupos de pacientes<sup>42</sup>. El oportunismo y los errores a los que puede llevar este tipo de análisis hacen necesaria su especial consideración<sup>42-45</sup>.

Se han publicado y son conocidos listados amplios de errores cometidos cuando se considera la respuesta por separado en subgrupos del estudio. Muchas veces se dan equivocadamente (e intencionadamente) por probados determinados beneficios terapéuticos en algunos subgrupos de pacientes, en base a los análisis erróneos cometidos en la valoración de los resultados de estudios en dichos subgrupos. Algunos de estos errores son ahora considerados obvios, pero en su momento fueron aceptados como proposiciones ciertas: así se consideró erróneamente probado que la aspirina era ineficaz en la prevención secundaria del ictus en mujeres, que el tratamiento antihipertensivo era ineficaz en prevención primaria de eventos cardiovasculares en mujeres, que los beta-bloqueantes no son eficaces en pacientes ancianos con infarto agudo de miocardio, que la trombolisis no es eficaz, o incluso es peligrosa, en el tratamiento del infarto agudo de miocardio en pacientes con infarto de miocardio previo, o que tamoxifeno no es eficaz en el tratamiento del cáncer de mama en mujeres menores de 50 años<sup>42</sup>. Estas conclusiones de ineficacia en subgrupos de pacientes son indudablemente graves, pero con suerte frecuentemente son fácilmente desechadas cuando otros estudios demuestran su error. Es más difícil que suceda lo opuesto: que una falsa conclusión sobre la aparente *eficacia* de un tratamiento en un subgrupo de pacientes sea descartada por posteriores estudios; y estos falsos principios aparentemente probados se perpetúan en la literatura médica, siendo difícil rebatirlos cuando se instalan en la práctica habitual.

El interés del análisis de subgrupos proviene de la necesidad del médico asistencial de saber lo que es mejor para un paciente concreto, cuando este paciente, por sus características, pertenecería a un subgrupo de los pacientes del estudio. Para algunos autores, éste análisis presenta un interés legítimo si ese subgrupo de pacientes (todos los cuales cumplen con los criterios de inclusión y exclusión) presenta claras diferencias en la evolución de las variables del estudio, con o sin tratamiento, respecto al resto del grupo; o si las diferencias en el proceso fisiopatológico entre ese subgrupo y el resto es importante y puede influir en la respuesta; o si existen dudas respecto al beneficio del tratamiento en ese subgrupo; o, finalmente, si en la práctica asistencial dicho subgrupo está infratratado<sup>42</sup>. En todo caso los análisis "post hoc" -análisis de los datos ya existentes, con el fin de profundizar en la información disponible y maximizar la rentabilidad de un estudio en un momento dado- son muy habituales para realizar éstos análisis de subgrupos. No obstante, no por frecuentes, el traslado de sus conclusiones a la práctica clínica real debe ser inmediata y sin una reflexión previa. Ya que la

robustez de estos análisis, tanto metodológica como científicamente y el objetivo/interés con el que han sido realizados, no es equiparable a la de los ensayos clínicos originales<sup>46</sup>.

Es llamativo que, siendo los epidemiólogos y estadísticos los que más han alertado sobre el riesgo del análisis de subgrupos, a veces ellos mismos presenten opciones o alternativas metodológicas de análisis de los datos para dar mayor relevancia y fiabilidad a dicho análisis: estadística bayesiana de probabilidad a posteriori frente al análisis frecuentista, test de interacción entre subgrupos y el efecto del tratamiento<sup>42,46</sup>. Es cierto que metodológicamente éstos análisis suponen un avance, y más adelante se expondrán mejor algunos, pero aún deben demostrar que el riesgo de error disminuye con su uso.

Para muchos autores el análisis de subgrupos es siempre discutible, no deben ser considerados nunca relevantes salvo que los datos de un estudio sean replicados por otros estudios diferentes, y sus conclusiones nunca deben ser tenidas en cuenta más que como generadoras de nuevas hipótesis, excepto que<sup>43,44</sup>:

- a. Esos subgrupos hayan sido tenidos en cuenta en la selección de pacientes
- b. Existan datos que indiquen un pronóstico diferente en la evolución de la enfermedad entre los subgrupos (diferencias de riesgo) y/o de diferencias en la respuesta al medicamento
- c. En base a estas diferencias:
  - Se haya llevado a cabo una aleatorización estratificada de los pacientes en el proceso de asignación del tratamiento.
  - Se compruebe, al finalizar el estudio, que las características basales de los subgrupos para el análisis de la eficacia (pacientes de ese subgrupo asignados al tratamiento experimental y asignados al tratamiento control) no muestren diferencias estadísticamente significativas entre sí.
  - El plan de análisis estadístico esté inicialmente planificado, y no se trate de una decisión “post hoc”.

En relación con el análisis y la replicación de los datos obtenidos en subgrupos, un ejemplo demostrativo de lo intrincado de estas valoraciones es el siguiente: en 1996 un ensayo clínico controlado, aleatorizado y doble ciego en pacientes con insuficiencia cardíaca crónica grave, ponía a prueba la hipótesis de que amlodipino, a diferencia de otros calcio-antagonistas, podría reducir un 25% respecto a placebo una variable combinada de morbilidad y mortalidad cardiovascular<sup>47</sup>. Este objetivo se evaluó también en los subgrupos de pacientes con cardiomiopatía isquémica y no isquémica, que se habían tenido en cuenta en el diseño del estudio para realizar una aleatorización estratificada por esta variable, ya que datos previos indicaban que los pacientes con cardiopatía isquémica podrían beneficiarse más de la administración de amlodipino<sup>47</sup>. En contra de lo esperado, ni en el total de 1153 pacientes, ni en los 732 pacientes con cardiomiopatía isquémica se encontró que amlodipino fuese más eficaz que placebo. Y, sorprendentemente, en contra de lo previsto, sí lo fue para el subgrupo de 421 pacientes con cardiomiopatía no isquémica<sup>47</sup>. Dado el resultado inesperado, los autores decidieron repetir el ensayo clínico, evaluando ahora como objetivo primario si amlodipino podría reducir la mortalidad, pero incluyendo en este segundo estudio solamente pacientes con cardiomiopatía no isquémica<sup>48</sup>. Los resultados de este ensayo

han sido negativos, y, en palabras de los autores, “estos resultados indican la necesidad de tener mucha precaución cuando se observan beneficios llamativos en subgrupos de pacientes, en ensayos clínicos que no se han diseñado con el objetivo primario de evaluar tales beneficios”<sup>48</sup>.

La inflación de datos falsos positivos como resultado de la realización de comparaciones estadísticas múltiples abarca tanto a los análisis pre-especificados de subgrupos como a análisis post hoc. La diferencia relevante entre ambos es que en el primer caso el número de comparaciones es conocido, pueden exigirse y establecerse las correcciones necesarias (por ejemplo, de Bonferroni), para establecer si la diferencia encontrada es o no estadísticamente significativa. Por el contrario, no es posible saber cuántas comparaciones no pre-especificadas han realizado los autores de la publicación, y cómo, hasta “encontrar” un efecto llamativo en un subgrupo de pacientes, y por lo tanto si se han aplicado o se han escamoteado las correcciones estadísticas necesarias a los límites de error alfa<sup>49</sup>. Los resultados negativos de análisis de subgrupos en ocasiones son también utilizados de modo maliciosos para plantear dudas no razonables y confusión sobre los resultados positivos de otros ensayos bien realizados<sup>49</sup>.

Más allá de los estándares para el análisis de subgrupos en los ensayos clínicos controlados, un punto absolutamente necesario, por su trascendencia y repercusiones, es resolver la falta absoluta de criterio para el análisis de los objetivos secundarios en las revisiones sistemáticas y los metaanálisis<sup>50</sup>. El riesgo de errores y equivocaciones de elevada trascendencia práctica de estas recomendaciones, debido al alto impacto de estos análisis, requiere medidas urgentes y bien definidas.

**¿Un paso adelante?  
Medicina basada  
en la evidencia,  
metaanálisis,  
revisiones  
sistemáticas y guías  
terapéuticas**

Hemos comprobado que los ensayos clínicos cumplen a la perfección con el objetivo para el que fueron diseñados, pero que su validez externa es limitada y no tiene la misma fiabilidad que su validez interna. Los médicos asistenciales están profundamente preocupados por esta limitación, que se extiende a las dudas que manifiestan sobre la fiabilidad de las revisiones y guías basadas en los ensayos clínicos<sup>35</sup>. Existen pruebas suficientes para concluir que hay un claro distanciamiento entre la práctica médica habitual y las recomendaciones expuestas en las diversas y múltiples guías existentes, lo que a veces origina una infrautilización de tratamientos que han demostrado fehacientemente su beneficio clínico, y otras veces la utilización excesiva de tratamientos que no se admiten en las recomendaciones terapéuticas publicadas<sup>51-53</sup>. Las dudas sobre la validez externa de los ensayos clínicos podría ser una de las explicaciones más razonables que justificaría ese comportamiento en la práctica asistencial<sup>35</sup>.

Admitamos que la evaluación de la validez externa es compleja, que requiere conocimientos de metodología estadística pero sobre todo una alta cualificación clínica, y que el objetivo de los ensayos clínicos de los nuevos medicamentos es cuantificar el beneficio de una intervención terapéutica en condiciones experimentales, y no cuantificar el beneficio resultante de su uso en la práctica asistencial. Y admitamos que se ha generado una enorme confusión derivada de la confrontación entre el juicio clínico de los médicos asistenciales y los análisis estadísticos que dan lugar a recomendaciones y guías terapéuticas. Por una parte es vital que el tratamiento administrado a los pacientes se utilice del modo más correcto posible en la mayor parte posible de casos. Y por otro que las recomen-

daciones tengan en cuenta la compleja situación asistencial. ¿Está justificado que sea impopular para muchos médicos asistenciales la imposición de guías terapéuticas basadas en criterios de autoridad, y esgrimidas por gestores asistenciales, responsables de política asistencial o grupos académicos, y respaldadas frecuentemente por la industria farmacéutica? ¿Es más razonable dejar la decisión en manos de los médicos asistenciales, sin que existan recomendaciones obligatorias? ¿Cómo se ha llegado a esta situación de desavenencia y críticas mutuas?

**Medicina basada en pruebas (Evidence-Based Medicine; en adelante: MBE)**

En 1992 se presentó formalmente la MBE por un grupo de trabajo, denominado Evidence-Based-Medicine Working Group<sup>54</sup>. Este grupo proponía la MBE como un nuevo paradigma científico. Paradigma significa ejemplo o modelo<sup>4</sup>, y se emplea la expresión *paradigma científico* tal y como la describió Kuhn en 1962, para referirse al conjunto de prácticas que definen una disciplina científica durante un período de la historia específico<sup>55</sup>. Un paradigma científico se presenta como un modo de ver y compilar los avances científicos que define tanto los problemas que pueden abordarse como la gama de pruebas admisibles que pueden llevar a su solución<sup>55</sup>. Cuando los defectos en un paradigma previo existente se acumulan hasta tal punto que el paradigma ya no es sostenible, es sustituido por una nueva forma de mirar el mundo. La ciencia médica estaba cambiando en la segunda mitad del siglo XX, y cambiando la práctica médica, fundamentalmente por la evolución de la investigación clínica tras la introducción de la metodología científica de los ensayos clínicos. Estos cambios, que implicaban por una parte que no se podía aceptar ningún nuevo medicamento que no hubiese demostrado previamente su eficacia mediante ensayos clínicos, y por otra parte requerían el uso más eficaz de la literatura médica para guiar la práctica médica asistencial, se presentaban por el grupo como un cambio de paradigma<sup>54</sup>.

Es extraordinariamente relevante al leer la presentación de la MBE, observar qué sentido y finalidad dan al nuevo paradigma los autores. Ya se puede leer en el subtítulo que se trata de una *nueva aproximación a la enseñanza de la práctica de la medicina*, y utilizan para ilustrar el paradigma un escenario clínico apropiado: un residente junior que se enfrenta a un caso clínico. En el pasado, con el viejo paradigma, el residente recurriría a su residente mayor, a la autoridad, para resolver el caso. Ahora, con el nuevo paradigma, el residente se va a la biblioteca, y resuelve *él mismo* la situación, leyendo la literatura científica. La aceptación de la MBE requiere entender que por una parte se precisa ser un médico competente con habilidades clínicas bien desarrolladas, que conocer la fisiopatología de las enfermedades es necesario, pero no suficiente para guiar la clínica asistencial, y que hay que conocer bien ciertas reglas para comprender e interpretar correctamente la literatura científica médica. El nuevo paradigma pone mucho menos énfasis en el valor de la autoridad, y mucho más en el desarrollo de habilidades que permitan a cada médico por una parte tomar decisiones sobre las pruebas existentes, y por otra evaluar la credibilidad y las opiniones ofrecidas por los expertos<sup>54,56</sup>.

Se acepta de modo amplio que la MBE ha cambiado sustancialmente el modo de ver y practicar la medicina, y que ha supuesto un avance notable en este sentido. Se han alcanzado logros como la Cochrane Collaboration para recoger y resumir las pruebas de los ensayos clínicos; se ha avanzado en los estándares metodológicos y de publicación de ensayos clínicos, se han puesto en marcha estructuras nacionales e internacionales para la elaboración de guías terapéuticas, etc. Y hay ejemplos incuestionables de logros obtenidos, desde muy pronto,

al incorporar criterios de MBE: en 1990 se elaboró una guía de tratamiento del asma, basada en ensayos clínicos y en estudios observacionales<sup>57</sup>. Se impulsó un cambio notable en los medicamentos utilizados, sobre todo en corticoesteroides inhalados, aunque los cambios en mortalidad no se pudieron concluir con fiabilidad<sup>58</sup>. Los promotores del nuevo paradigma han insistido una y otra vez en que se trata de integrar la experiencia clínica y las pruebas publicadas, y han redefinido posteriormente la MBE como *el uso consciente, explícito y juicioso de las mejores pruebas disponibles en la toma de decisiones sobre el cuidado de los pacientes individuales*. Recuerdan, además, que la práctica de la MBE requiere integrar la experiencia asistencial individual con las mejores pruebas externas disponibles en una investigación sistemática<sup>59</sup>. En este último artículo de opinión, además de redefinir la MBE los autores tratan de responder a las críticas planteadas a la MBE: señalan que las dificultades para llevarla a la práctica asistencial no son insalvables, que no es un proceso para que sea elaborado desde torres de marfil o desde sillones académicos, y no es un libro de recetas de cocina<sup>59</sup>.

Pero es extraordinariamente llamativo cómo se ha retorcido y se ha mistificado el concepto original de MBE. Transformando el propósito original de la MBE, que era la educación y la formación de los médicos, es muy frecuente que se utilice este paradigma como argumento de autoridad por los que defienden la autoridad, yendo en contra de opiniones individuales razonadas que por otra parte pueden estar bien fundamentadas. Es decir, en vez de utilizar esta metodología de trabajo para educar a los médicos con el fin de que tomen sus propias decisiones, se les confunde y agobia, basándose en un erróneo concepto de MBE, o en un concepto mal entendido, o maliciosamente utilizado para eliminar las críticas y las dudas sobre las opiniones de pretendidos expertos, teniendo así que aceptar estas opiniones un médico asistencial bien formado, que se encuentra *abrumado por la eminencia, y no por la evidencia*, a pesar de que su análisis personal de la literatura le lleve a concluir que lo beneficioso para su paciente es distinto a lo que indican esos expertos.

Ahora todo es MBE, un concepto cuyo uso ha sido y es extraordinariamente creciente. Una de las causas de la deriva en la utilización de la MBE es que, en contra de la propuesta del grupo creador, que recalca que la MBE necesita recurrir en muchas ocasiones a estudios observacionales que hayan evaluado lo que ocurre en la práctica asistencial<sup>59</sup>, se considera por muchos autores como un proceso que se basa exclusivamente en los datos derivados de ensayos clínicos controlados. El motivo de este modo de presentar la MBE es que así resalta su supuesta base científica y la autoridad de sus juicios, olvidando los errores de extrapolación de los datos, y sirve en ocasiones para una más fácil instrumentalización potencial de la gestión y asignación de recursos asistenciales, la llamada Atención Sanitaria Basada en Pruebas (Evidence-based Health Care), o para implementar la Política de Sanidad Basada en Pruebas (Evidence-based Health Policy)<sup>60</sup>.

Se plantea en la actualidad que la MBE puede estar en crisis. Muchos autores tratan de limitar el fervor casi religioso que ha suscitado la MBE, llamando la atención sobre todo a las dificultades para extrapolar los datos de los ensayos clínicos a la Atención Primaria, debido principalmente, como se ha indicado antes, a los criterios de selección de la población en estudio en los ensayos clínicos<sup>61</sup>. Según estos autores, y grupos como el autodenominado Evidence Based Medicine Renaissance Group, algunos de los problemas que pueden haber conducido a esta crisis son los siguientes<sup>61-63</sup>:

- Las dificultades prácticas y económicas de aplicar a los pacientes sólo las intervenciones que presenten alto valor según los principios de la MBE. Dificultades que comprometen la práctica de la denominada Medicina Clínica Basada en las Personas, o la Práctica Clínica Basada en las Pruebas, que debe tener en cuenta no sólo las pruebas disponibles, sino también la opinión del paciente. José Antonio Sacristán en Patient-centered medicine and patient-oriented research: improving health outcomes for individual patients. BMC Med Inform Decis Mak. 2013;13:6. doi: 10.1186/1472-6947-13-6, sostiene que “el objetivo final de la Medicina Basada en las Personas, o Medicina Basada en el Paciente, forma parte de la tendencia a una mayor personalización de la medicina, dando más peso al papel de los pacientes en los sistemas de salud, y tiene como objetivo no el conocimiento de cuál es el mejor tratamiento para un paciente tipo, sino mejorar los beneficios de salud de cada paciente individual, asegurándose de que los resultados de la investigación médica tienen un impacto real en la atención sanitaria. No debe considerarse una aproximación contrapuesta a la Medicina Basada en las Pruebas, sino una aproximación complementaria, facilitadora de la convergencia entre la investigación y la práctica asistencial”
- La deformación de la marca “evidence-based”. Como marca, no como nuevo paradigma para una práctica clínica sustentada en pruebas científicas y contando con la opinión del paciente, alguna compañía farmacéutica estableció en su momento su estrategia de marketing.
  - Estos agentes definen qué es una enfermedad, seleccionan cuáles son los factores de riesgo de interés, y eligen cuáles son las variables subrogadas que les interesa para establecer la eficacia de una intervención terapéutica.
  - Realizan estudios con un número elevado de pacientes, con el fin de disponer de un poder estadístico sobredimensionado para asegurar que diferencias muy pequeñas alcancen significación estadística, con una selección de la población del estudio adecuada a sus intereses.
  - Realizan estudios destinados a obtener avances marginales, sobreestimando potenciales beneficios e infraestimando riesgos.
  - Se publican sobre todo estudios con resultados positivos, y no los que presentan resultados negativos, lo que limita el acceso a las pruebas reales de la eficacia de los medicamentos o productos sanitarios.
- A medida que el lenguaje de la MBE se integra más y más en la práctica médica, se burocratizan reglas y algoritmos que se convierten en formas aceptadas (por la autoridad de quien los establece) de aplicar *la mejor evidencia disponible*. Las mejores pruebas son silenciadas a favor del énfasis en las reglas y el control tecnocrático. La evaluación de riesgos utilizando puntuaciones y algoritmos “basados en la evidencia”, se produce a escala industrial, alejada de las decisiones médicas personalizadas.

Uno de los aspectos más negativos de la aplicación “industrial” de la MBE, lo expresan algunos autores con la expresión “too much evidence”, que irónicamente, constituye una medida del éxito del paradigma<sup>61,63</sup>. Con esta expresión se hace referencia a la enorme cantidad de publicaciones y datos de MBE disponibles. El número de directrices publicadas hace que sean insondables y difíciles de manejar. Comentan un ejemplo: una auditoría de la actividad médica que se lleva a

cabo durante 24 horas en un hospital de agudos señalaba que en ese tiempo se evaluaron 18 pacientes con 44 diagnósticos, para los que se identificaron 13679 páginas de guías nacionales relevantes para la atención de esos pacientes (que requerirían 122 horas de lectura)<sup>63</sup>!

Las revisiones sistematizadas de la información disponible, de los datos probados, y las guías resultantes, constituyen en estos momentos un elemento determinante para considerar eficaz, e incluso efectiva, una medida terapéutica determinada. La saturación de guías hace necesaria una revisión de su validez y objetividad. Si, además de la imposibilidad de abarcarlas, la información que ofrecen no fuese la pertinente para que el médico pueda decidir el uso racional de los medicamentos, según establecía la OMS<sup>5</sup>, no se habría cumplido con el objetivo deseado, sino que, por el contrario, paradójicamente nos habríamos alejado de él.

Hay que ser cuidadosos en relación con la información no adecuada, y con la idea señalada antes de la utilización orientada al marketing de la marca “evidence-based”. Hay que dilucidar siempre los conflictos de interés debidos a la financiación por las compañías no sólo de la investigación científica, sino de los análisis de los datos disponibles para la elaboración de revisiones sistemáticas que ha quedado, afortunadamente, balanceada por el peso y la importancia que le damos a los trabajos de los Centros Cochrane. Los conflictos de interés han sido objeto de atención desde la década de 1980, y han sido definidos como un conjunto de condiciones en las que el juicio profesional en relación con un interés primario, el bienestar del o la validez de la investigación, tiende a ser indebidamente influenciado por un interés secundario más espúreo<sup>64</sup>. En un artículo publicado hace algunos años, se evaluaron 37 estudios cuyo objetivo fue evaluar el alcance, impacto, o la gestión de la relación financiera entre la industria, los investigadores o las instituciones académicas<sup>65</sup>. La conclusión de los autores fue que existía una asociación relevante y estadísticamente significativa entre la financiación de la industria y las conclusiones favorables a los intereses de la industria de esos estudios financiados<sup>65</sup>. En una publicación más reciente, los autores abordan los conflictos de interés financieros relacionados con los resultados no de estudios individuales, sino de las revisiones sistemáticas; para ello toman como ejemplo las revisiones sistemáticas relacionadas con estudios que evalúan la asociación entre el consumo de bebidas azucaradas y el aumento de peso<sup>66</sup>. Los autores concluyen que claramente los potenciales conflictos de interés relacionados con la financiación de los estudios influyen en los resultados de esas revisiones, y ofrecen el siguiente dato: entre las publicaciones que no manifiestan conflictos de interés, más del 83% concluyen lo contrario, es decir, que el consumo de bebidas azucaradas podría considerarse un factor de riesgo de ganancia de peso; cuando las revisiones manifiestan conflictos de interés financieros, más del 83% concluye que no hay pruebas suficientes que permitan sostener que existe una asociación entre el consumo de bebidas azucaradas y la ganancia de peso<sup>66</sup>.

La sobreproducción de revisiones sistemáticas ha conducido a una situación más compleja, menos abarcable aún, y más manipulable: las *overview of reviews* o visiones de conjunto de revisiones. Estas aproximaciones pretenden digerir y facilitar el resultado de distintas revisiones sobre un tema concreto, y se considera un modo pragmático, fiable y novedoso de *síntesis de la evidencia*, aportando una visión de conjunto de todas las pruebas existentes, para extraer conclusiones sobre un tratamiento o abordaje terapéutico, ante la sobreproducción de información existente, entendiendo que el exceso de revisiones puede confundir al médico

asistencial. Un modo sofisticado de aportar información útil para tomar decisiones en MBE. En un reciente artículo se ha llevado a cabo un análisis descriptivo de 75 de estas visiones de conjunto de revisiones, publicadas entre los años 2000 y 2011, que cumplían criterios que se habían establecido previamente para ser tenidas en cuenta: debían tener un objetivo definido, evaluar una intervención concreta (farmacológica, no farmacológica o mixta), utilizar métodos bien especificados para identificar cómo se habían seleccionado las revisiones sistemáticas objeto de análisis, cómo se habían recogido y sintetizado los datos de estas revisiones, y debían utilizar en su elaboración sólo revisiones<sup>67</sup>. En sus conclusiones, los autores indican que el principal problema registrado es la variabilidad en la aproximación metodológica utilizada en el análisis de conjunto. Indican que la utilidad de estos análisis depende en gran parte de la calidad y disponibilidad de las revisiones utilizadas, y en último término de los ensayos clínicos u otros estudios clínicos utilizados para la elaboración de las revisiones; las deficiencias metodológicas en cada nivel (ensayos, revisiones) comprometen los resultados y las conclusiones del análisis de conjunto, y en consecuencia condicionan su validez, su fiabilidad y su utilidad para la toma de decisiones<sup>67</sup>. Así pues, la situación es compleja, poco abarcable, y de fiabilidad dudosa. Otro estudio similar, que evalúa 126 visiones de conjunto de revisiones sistemáticas, publicadas en un período de tiempo similar, llega a la misma conclusión: estas visiones de conjunto de revisiones sistemáticas a menudo carecen del rigor metodológico necesario<sup>68</sup>. A pesar de todo, y debido a la avalancha abrumadora de la información publicada y al recurso constante, e insistente (aunque subrepticio) a la autoridad, las revisiones sistemáticas y las visiones de conjunto de las revisiones sistemáticas ejercen una influencia indudable en la MBE y en la toma de decisiones asistenciales.

***Las Guías  
Terapéuticas basadas  
en las pruebas  
(Evidence-Based  
Guidelines; en  
adelante GTE) y el  
nivel de evidencia***

En la década de 1980 comenzó a implantarse de modo generalizado el uso de Guías Terapéuticas como un modo de organizar y sistematizar la asistencia médica. Su desarrollo fue paralelo al de la MBE, pero su objetivo y finalidad eran diferentes, al menos en aquel momento. El desarrollo de las Guías se presentó como una respuesta a una demanda de los médicos, con el fin de tratar de conseguir algunas de las finalidades del programa CER: reducir la atención asistencial inadecuada, disminuir las variaciones geográficas en los patrones de práctica, y hacer un uso más eficaz de los recursos sanitarios<sup>69</sup>. Muchos gobiernos como el de USA crearon nuevas agencias con la responsabilidad de elaborar guías asistenciales, ofreciendo su desarrollo a distintas organizaciones de médicos y sociedades científicas especializadas. También se involucraron centros médicos académicos formando consorcios de investigación en las GTE. Además se incorporaron otros agentes, como los hospitales, las aseguradoras, o las empresas privadas<sup>69</sup>. A pesar de la multiplicidad de intereses, de su variable complejidad y fiabilidad, incluso de los errores en la percepción del significado de las guías (en alguna encuesta, un 7% de los médicos encuestados se declaraban familiarizados con guías ficticias sobre las que socarronamente se preguntaba en la encuesta), para muchos las GTE eran un paso necesario regido por el sentido común: consideraban necesario “guiar” a los médicos<sup>70</sup>. Pronto, con el auge de la MBE, se presentó el desarrollo de las guías como la elaboración de GTE, en base al consenso de distintos participantes que tratan de aunar la prestación de una actividad asistencial óptima (mezcla de las pruebas científicas de los estudios clínicos y de la opinión de expertos), los intereses de la gestión sanitaria, y la discusión por expertos del protocolo final<sup>71</sup>.

Lo más llamativo de todo esto es que, con este procedimiento, pretendían que estas recomendaciones estuviesen directamente vinculadas con la *evidencia científica de efectividad*, asumiendo que se pueden extrapolar fácilmente los datos de eficacia a previsiones de efectividad<sup>71</sup>. En algunos países, como Inglaterra, se ha creado el National Institute for Health and Clinical Excellence (NICE), cuyas guías gozan de un enorme respeto profesional<sup>72</sup>.

Desde luego, las GTE nos han alejado del juicio crítico profesional personalizado, sustituyéndolo por argumentos de autoridad. Y nos plantean dos problemas nuevos. En primer lugar, no sólo elaboran GTE los organismos oficiales: se generan múltiples guías locales que cubren la rutina asistencial y las situaciones de riesgo. Se ha publicado que el número de guías y normativas disponibles en la intranet de tres servicios de agudos del National Health Service de Londres oscila entre 192 y 457<sup>73</sup>. Además, el equipo médico tiene que entender y cumplir con guías y normas de gobierno locales y profesionales. Los autores de la publicación señalan que si un paciente es admitido para una intervención quirúrgica de urgencia por una fractura de cuello de fémur, las diferentes fases de su atención asistencial están regidas por 75 guías y normas<sup>73</sup>. Para los autores del artículo, la proliferación de guías, la multiplicidad de normas sobre el mismo tema, la accesibilidad difícil y la complicada lectura de las guías, extensas, complejas y a veces mal redactadas, hacen que sea prácticamente imposible conocerlas todas, y cumplir con ellas, y, peor aún, que sea sumamente difícil distinguir lo esencial de lo irrelevante o erróneo, reduciendo así su cumplimiento, y con un resultado impredecible en la calidad asistencial<sup>73</sup>.

Se plantea también un segundo problema, que incide aún más directamente en el tema de esta monografía. Lejos de su utilidad para que el médico realice reflexiones como las sugeridas inicialmente por la MBE, la guía le *impone* al médico una conducta, esté o no de acuerdo en su racionalidad, y le *exige* que tenga confianza en que los que han generado la guía saben lo que hacen. Las guías realizan análisis de los datos de eficacia, y, mediante modelizaciones o mediante razonamientos de consenso *establecen normas de efectividad comparada*: le indican al médico qué medicamento debe usar como primera o segunda elección, y, sin criticar la validez externa de los estudios analizados, supone que la respuesta en la práctica asistencial, la efectividad, será la misma que la eficacia detectada en los ensayos clínicos. Pero, dejando aparte por ahora los sesgos de extrapolación, ¿al menos se evalúan correctamente los datos de los ensayos clínicos y de los estudios observacionales en las guías? Por su relevancia, pongamos algunos ejemplos de que quizás se deba dudar de esa sabiduría que se supone a los autores de las guías.

### **Caso 1. Ausencia de datos. Adrenalina en resucitación cardiopulmonar**

El European Resuscitation Council mantiene un ciclo de actualización de su prestigiosa Guía de Resucitación Cardiopulmonar cada 5 años. La última actualización se presentó en 2010<sup>74</sup>. En la sección 4, donde se establecen las guías para soporte vital avanzado en adultos, se indica que a pesar del uso generalizado de adrenalina, y en algunos casos de vasopresina, durante la reanimación cardiopulmonar, no hay estudios controlados con placebo que demuestren que los vasopresores aumenten la supervivencia al alta hospitalaria en estos pacientes, aunque mejoren la supervivencia a corto plazo<sup>75</sup>. Y continúa argumentando que los participantes en la conferencia de consenso han debatido a fondo las recomendaciones. A pesar

de la ausencia de datos que demuestren aumento de la supervivencia a largo plazo, adrenalina se considera el vasopresor estándar en la parada cardíaca. La práctica actual apoya esta alternativa, *aun en ausencia de pruebas*<sup>75</sup>.

Para muchos autores, en ese momento, sí había suficientes datos que sugerían que adrenalina estaba lejos de ser el agente vasopresor ideal para estos pacientes, y el uso de adrenalina no cumplía los criterios de la MBE<sup>76</sup>. La opinión de estos críticos se vio reforzada cuando en 2011 se publicó el primer y hasta ahora único ensayo clínico controlado con placebo de adrenalina en pacientes en parada cardíaca. El objetivo primario del estudio fue la supervivencia al alta hospitalaria, y el ensayo no demostró diferencias significativas entre ambos tratamientos<sup>77</sup>. Apoyando esta conclusión, se ha publicado un estudio observacional con una de las mayores bases de datos de resucitación cardiopulmonar, donde se concluye que el uso de adrenalina podría estar asociado con una menor supervivencia y con un empeoramiento de las variables neurológicas de estos pacientes<sup>78</sup>. La guía no ha sido modificada.

El problema ha sido la elaboración de una guía no basada en datos válidos, sino en criterios de autoridad. Las alternativas, son: ¿Se propone el seguimiento de la guía, basado en la autoridad de los redactores? ¿Se descartan los resultados del único ensayo clínico existente y del estudio observacional que lo respalda? ¿Es posible la realización de un ensayo clínico, con las dificultades éticas de llevarlo a la práctica, ya que las condiciones de inclusión en el estudio (parada cardíaca) son difíciles, y existe ya una guía ya publicada y de prestigio que establece cuál debe ser el comportamiento médico? Las dificultades éticas para llevar a cabo el ensayo clínico controlado, aleatorizado y doble ciego PARAMEDIC2, que aborda esta cuestión, ilustran este conflicto<sup>79</sup>.

### ***Caso 2. Distorsión de los datos. Tratamiento de la hipertensión arterial en el anciano***

Las recomendaciones sobre el tratamiento de la hipertensión arterial en mayores de 80 años son controvertidas, con pruebas no concluyentes sobre el beneficio potencial de tratar a estos pacientes; los estudios epidemiológicos han sugerido una y otra vez que la presión arterial y el riesgo de muerte estaban inversamente relacionados en estos pacientes, de tal modo que la reducción de la presión arterial podría incrementar el riesgo de muerte; por lo que era muy importante contar con un ensayo clínico que pudiese aportar datos fidedignos en este campo. El estudio HYVET se planificó con este fin<sup>80</sup>.

El ensayo clínico HYVET se diseñó para evaluar la eficacia de una pauta de indapamida y perindopril (ajustada para intentar alcanzar una presión arterial objetivo de 150/80 mmHg) vs placebo, con un diseño aleatorizado y doble ciego, que incluía pacientes de al menos 80 años de edad, de ambos sexos; se consideraban hipertensos si presentaban unos valores medios (media de 4 medidas separadas un mes) de presión arterial sistólica en sedestación de 160-199 mmHg, tras un tiempo sin tratamiento activo (sólo con placebo), y una presión arterial sistólica en bipedestación de al menos 140 mmHg<sup>80</sup>. A mitad del estudio se suspendió el requisito de que el valor medio de su presión arterial diastólica en sedestación fuese de 90 a 109 mmHg, pudiendo incluir pacientes sólo con presión arterial sistólica aislada. El objetivo primario del estudio fue evaluar la incidencia de cualquier ictus (mortal o no), sin considerar los ataques isquémicos transitorios;

entre los objetivos secundarios estaba la muerte, por cualquier causa o de origen cardiovascular<sup>80</sup>. Los resultados del estudio fueron que el tratamiento activo no consiguió una reducción estadísticamente significativa en la tasa de incidencia de ictus (objetivo primario). Respecto a los objetivos secundarios, tampoco se redujo significativamente la muerte por ictus o por causas cardiovasculares, aunque sí disminuyó la muerte por cualquier causa y la tasa de insuficiencia cardíaca. En el caso de que la reducción en los ictus hubiese alcanzado significación estadística, las conclusiones indican que habría que tratar a 188 pacientes durante un año para que uno de ellos, que habría presentado un ictus sin el tratamiento, no lo sufriese. Los otros 187 pacientes presentarían un ictus, o no, independientemente del tratamiento<sup>80</sup>.

La Guía de 2013 de la European Society of Hypertension and of the European Society of Cardiology para el manejo de la hipertensión arterial presentaba así su versión de estos hechos: al elaborar la guía de 2007 de estas sociedades no se disponía de evidencia directa sobre los efectos del tratamiento antihipertensivo en ancianos con HTA (mayores de 80 años). La posterior publicación de los resultados del estudio HYVET, comunicó una reducción significativa de las complicaciones cardíacas graves y la mortalidad por todas las causas asociada a unos objetivos de PA < 150 mmHg en octogenarios con una presión arterial sistólica superior a 160 mmHg<sup>81</sup>.

Como puede observarse, en la guía no se indica que los valores de la presión arterial inicial se han recogido en sedestación, que en bipedestación los valores de presión arterial sistólica debían ser superiores a 140 mmHg, que los beneficios señalados no se refieren a los objetivos primarios (y por lo tanto científicamente no pueden establecerse conclusiones válidas respecto a ellos), ni se comentan los resultados del objetivo primario o de otros objetivos secundarios, ni se comentan otros aspectos relevantes del estudio. La Guía para el médico está totalmente clara para los autores. Aunque se haya realizado manipulando, intencionadamente o no, la presentación de los datos disponibles. Y olvidando los datos de los estudios observacionales.

### ***Caso 3. Cuando las guías presentan como adecuados los objetivos terapéuticos subrogados***

En noviembre de 2013, un grupo de trabajo conjunto del Colegio Americano de Cardiología y la Asociación Americana del Corazón lanzó nuevas directrices en el tratamiento de pacientes con hipercolesterolemia, para reducir el riesgo de enfermedad cardiovascular aterosclerótica en adultos<sup>82</sup>. Este documento llegaba después de varios años de intensa deliberación, 12 años después de la tercera guía del Panel de Tratamiento de Adultos con hipercolesterolemia (ATP III), y 8 años después de una actualización de la guía ATP III recomendando la reducción más agresiva (a menos de 70 mg/dL) de los niveles de colesterol asociado a lipoproteínas de baja densidad (LDL-C) como una opción en pacientes de alto riesgo<sup>83,84</sup>. La nueva guía rompe radicalmente con una historia de recomendaciones que se remontan a la década de 1980, y para muchos autores constituye un cambio “valiente y sabio”, después de décadas de manipulación de datos asumiendo, de un modo carente de toda validez científica, que los cambios inducidos terapéuticamente en los niveles séricos de colesterol se asocian con una reducción del riesgo cardiovascular<sup>85</sup>. Ya en 1994, el primer estudio con simvastatina publicado

indicaba que, hasta ese momento, ningún tratamiento hipocolesterolemizante había podido demostrar una reducción de la mortalidad total o de la mortalidad cardiovascular, aunque redujesen el colesterol<sup>86</sup>. Con posterioridad, otros fármacos como ezetimiba<sup>87</sup>, niacina<sup>88</sup>, torcetrapib<sup>89</sup> o anacetrapib<sup>90</sup>, no han podido demostrar tampoco reducción del riesgo cardiovascular, a pesar de la reducción conseguida por esos medicamentos de la hipercolesterolemia o de LDL-C; en algunos casos el riesgo cardiovascular, tras el tratamiento farmacológico, es superior al de otras alternativas que controlan “peor” la hipercolesterolemia o el LDL-C.

La nueva guía de 2013 representa un cambio radical en el enfoque y la gestión de la hipercolesterolemia, y ha suscitado una controversia considerable, aún no finalizada. ¿El motivo?: la guía se centra solamente en las pruebas que indican que la intervención terapéutica consigue la reducción del riesgo cardiovascular. La reducción del riesgo cardiovascular es el objetivo terapéutico, y entiende que tomar como objetivo terapéutico subrogado los niveles de colesterol plasmático es un error que puede traer graves consecuencias para los pacientes. Este cambio implica que solamente deben ser utilizados los fármacos que han demostrado una reducción en el riesgo cardiovascular en pacientes con hipercolesterolemia (las estatinas), y que se debe abandonar el uso de otros medicamentos<sup>82</sup>. ¿Por qué se han impuesto las guías previas a los médicos asistenciales? ¿Con qué pruebas? ¿Alguien puede explicar a los médicos y las agencias empeñadas en el control del colesterol plasmático, que tomar como objetivo terapéutico esta variable subrogada ha sido probablemente un inaceptable error de medida de efectividad terapéutica? ¿Y que, probablemente, algunas de estas medidas incluso hayan aumentado el riesgo cardiovascular en los pacientes? ¿Hasta cuándo van a seguir vigentes las guías basadas en la ATP III?

#### ***Caso 4. Cuando las guías tienen baja calidad y conflictos de interés***

Desde 2011 se han realizado ensayos clínicos con más de 30 agentes directos antivirales para tratar la infección por el virus de la hepatitis C crónica (HCV). Entre estos agentes destaca sofosbuvir, ya que ha sido el primero de estos antivirales aprobado por la FDA para el tratamiento de los HCV de genotipos 1 a 4. En enero de 2014, the American Association for the Study of Liver Diseases y the Infectious Diseases Society of America elaboraron una guía para el tratamiento de la hepatitis C y recomendaron sofosbuvir para todos los pacientes, excepto aquellos con insuficiencia renal grave<sup>91</sup>.

En mayo de 2014 el Center for Evidence-based Policy de la Oregon Health and Science University emitió un informe para evaluar la guía antes citada, y la validez de los datos utilizados<sup>92</sup>. Respecto a los datos utilizados, el informe concluye que todos los ensayos utilizados para elaborar la guía presentan un elevado riesgo de sesgos, y ninguno podría ser considerado como un estudio con capacidad para ser aplicado a la práctica asistencial. En general, los estudios publicados se califican como mediocres. Solamente 1 de 10 estudios publicados utiliza un comparador que podría responder a la pregunta práctica siguiente: ¿las nuevas pautas terapéuticas con sofosbuvir ofrecen una mejor evolución clínica y un riesgo menor que los tratamientos estándar actuales<sup>92</sup>?

Además, el informe de la Universidad de Oregón evalúa la calidad de la guía publicada, y concluye que la calidad es muy baja, por dos motivos: el primero, porque no informa del riesgo de sesgos en los estudios utilizados. Y, el segundo,

porque los autores y los patrocinadores de la guía tienen *muchos y relevantes conflictos de interés*<sup>92</sup>. La International Antiviral Society-USA fue socio colaborador de la guía, y uno de los patrocinadores más importantes de esta sociedad es la industria farmacéutica, incluida la empresa Gilead. Como dato de interés, con posterioridad (octubre de 2014) la FDA autorizó a Gilead la comercialización de Harvoni (asociación de ledipasvir y sofosbuvir), el primer medicamento de combinación para el tratamiento de infecciones por HCV de genotipo 1<sup>93</sup>.

Otra reciente publicación cuantifica la baja calidad y los conflictos de interés en la elaboración de las guías terapéuticas; los autores justifican la realización del estudio por la, según ellos, utilización habitual de pruebas y estudios de baja calidad en la elaboración de las guías, y refieren otros estudios previos en cardiología que han concluido que la mayoría de los autores de las guías de cardiología presentan múltiples conflictos de interés<sup>94</sup>. Los autores de este Nuevo trabajo revisaron 149 guías elaboradas por the American Society of Diagnostic and Interventional Nephrology, the American Society for Gastrointestinal Endoscopy, the Society for Cardiovascular Angiography and Interventions y the American Thoracic Society and American College of Chest Physicians. Sus conclusiones son las siguientes: la mayor parte de las guías utilizan pruebas (evidence) de baja calidad, y presentan numerosos conflictos de interés<sup>94</sup>. Si las GTE son de baja calidad por los datos utilizados, y con frecuentes conflictos de interés, ¿cómo se puede obligar al médico asistencial a aceptar dichas guías, abrumándole con su extensión, complejidad, difícil lectura y criterios de autoridad para su cumplimiento?

El estudio citado, además, hace referencia a una ausencia frecuente de la *clasificación y jerarquización del grado de evidencia* de las recomendaciones<sup>94</sup>. Este es un aspecto con el que también se intimida y abruma al médico que lee la guía, ya que éste, en primer lugar, piensa que se trata de un sistema de clasificación bien evaluado y consensuado. Sin embargo, se trata de un sistema criticado reiteradamente por falta de solidez, y se considera por muchos inadecuado apoyarse en él para eliminar críticas o subrayar el valor de una recomendación. La clasificación jerarquizada del grado de evidencia de las recomendaciones se considera poco aplicable fuera del campo de la evaluación de los ensayos clínicos, a menudo se evalúan incorrectamente éstos en relación con la pregunta a contestar en la guía, no se tiene en cuenta la validez externa y aplicabilidad de los estudios, y no queda claro el valor operativo del sistema de clasificación jerarquizada del grado de evidencia<sup>95</sup>. Así, muchos grupos tratan de desarrollar sus propios sistemas para jerarquizar la fuerza de sus recomendaciones, criticando los anteriores<sup>95</sup>. Pero, si no se utilizan, ¿para qué sirven? ¿Aceptamos los sistemas cuyo uso se ha generalizado, aunque no hayan sido puestos a prueba de modo serio y hayan sido razonada y reiteradamente criticados? Rawlins, uno de los responsables del NICE, indica que en 30 años se han publicado más de 60 de estos sistemas de jerarquizar el grado de evidencia, que se han utilizado por agencias reguladoras, evaluadores de información, o grupos de elaboración de guías terapéuticas con el fin de valorar según su interés la calidad de las pruebas disponibles y establecer la fuerza de las recomendaciones que realizan<sup>96</sup>. Según este autor, el uso de estas jerarquías no sólo es defectuoso desde el punto de vista intelectual y científico, sino también arriesgado, y entiende que la adhesión servil a las jerarquías de evidencia pueden llevar a utilizar métodos inadecuados para contestar a preguntas concretas en la práctica asistencial en el mundo real<sup>96</sup>.



### III. El rol de la CER: decisiones, implicaciones y aportaciones

---

La brecha entre los datos de eficacia y las estimaciones de efectividad es, por tanto, considerable. Pongamos un ejemplo: se reivindica que medicamentos importantes como las estatinas, hipotensores o bifosfonatos son coste-efectivos. Este término debería significar que, en base a los datos de estudios experimentales, el coste relacionado con la administración de la intervención terapéutica (costes directos e indirectos) es aceptable, en relación con el beneficio obtenido tras su utilización en la práctica asistencial habitual (efectividad). Por lo general el término *coste* se utiliza erróneamente en éstas reivindicaciones: por una parte, no se tiene en cuenta el coste real de la intervención; para ello deberían considerarse los siguientes aspectos: consumo de recursos asistenciales, empleo de tiempo de los distintos profesionales sanitarios para su indicación/requisitos diagnósticos, coste de la intervención (en caso de un medicamento: producto + materiales y personas-tiempo implicadas en su prescripción, dispensación, preparación, administración y monitorización). Y por otra parte deben tenerse en cuenta los recursos empleados en el diagnóstico y tratamiento de las complicaciones o las reacciones adversas al medicamento (RAM), incluyendo de nuevo el tiempo de los profesionales sanitarios para controlar la RAM, y el coste de las intervenciones realizadas con este fin. A este valor hay que restar los gastos derivados de la falta de control de la enfermedad, y sumar o restar los recursos de salud consumidos durante los años adicionales de vida conseguidos. Algunas entidades, como el NICE, realizan unos informes en este sentido ejemplares, aunque su metodología reciba críticas en ocasiones<sup>97</sup>. Pero lo grave es que muy frecuentemente el análisis del coste, en nuestros medios, se simplifica de un modo burdo, grosero y falto de altura ética, política e intelectual cuantificando sencillamente como coste al valor de adquisición del producto.

Puesto que las repercusiones financieras de CER serán inevitables, y aunque no es el propósito de esta monografía, es conveniente entender cómo se definen algunos conceptos farmacoeconómicos en el programa CER y en otros medios<sup>98-100</sup>:

- *Coste*: la renuncia a beneficios alternativos que se hace cuando se utiliza un recurso para un propósito determinado. En economía de la salud, el coste se refiere a los recursos consumidos durante la prestación de la atención sanitaria, y describe la cantidad de dinero y recursos empleados en un determinado tratamiento o estrategia de gestión. El coste incluye, pero no se limita a, los costes médicos directos (por ejemplo, los costos asociados con los médicos, instalaciones, productos farmacéuticos, bienes duraderos), los costes no médicos directos (por ejemplo, los costes para los cuidadores, gastos de viaje), y los costes indirectos (otros gastos del producto interno bruto). Esta definición refleja los costes reales de la prestación de atención de salud y no incluye las reclamaciones, cargos u otros aspectos de la facturación de atención médica.
- *Beneficio*: medida cuantitativa de la mejora de la salud como resultado de los gastos realizados (coste), y típicamente se expresa en términos monetarios.

- *Valor*: medida del beneficio definido en términos de mejora objetiva de la calidad de vida (QOL)
- *Utilidad*: objetivo o expresión matemática de la preferencia por un determinado resultado o estado de salud, a menudo cuantificado utilizando un valor de QOL estándar
- *AVAC*: Años de vida ajustados por la Calidad de Vida (Quality Adjusted Life-Year; QALY). Es el producto de la utilidad por el período de tiempo durante el cual se crea este valor. Se trata de una medida universal de resultado de salud aplicable a todas las personas y todas las enfermedades que combina, en una sola medida, las ganancias o pérdidas en cantidad y en QOL.
- *Coste-utilidad*: la correlación (ratio) entre el coste y la utilidad describe un cierto tipo de relación entre el valor resultante de aplicar una intervención y el coste que la sociedad puede y está dispuesta a pagar para adquirir ese valor.
- *Análisis de coste-beneficio*: estudio de la relación monetaria entre los costes de atención de la salud y sus mejoras resultantes (donde los costes y los beneficios se expresan en unidades monetarias)
- *Análisis de coste-utilidad*: comparación entre opciones alternativas en términos de sus costes y resultados, donde los resultados se expresan en unidades de utilidad o preferencia (por ejemplo, QALY)

Mientras que la MBE principalmente trata de identificar la evidencia objetiva de la eficacia de distintas intervenciones, con estos análisis farmacoeconómicos se busca integrar los datos de eficacia clínica con los datos de los análisis de coste-utilidad para identificar estrategias médicamente eficaces para el manejo de enfermedades que también se caractericen por una relación coste-utilidad favorable.

No obstante, como se ha insistido antes, esta aproximación, incluso la realizada correctamente desde el punto de vista del coste, se lleva a cabo *asumiendo erróneamente que la efectividad*, en la práctica asistencial, será la misma que la eficacia previamente detectada en los ensayos clínicos. Y el estudio de la efectividad es necesario para realizar análisis de coste-efectividad: este tipo de análisis no es comparar opciones terapéuticamente similares para seleccionar la más barata, sino que consiste en analizar si los beneficios añadidos que se derivan del uso habitual de una opción terapéutica en condiciones reales asistenciales compensan su coste adicional. La finalidad última de este tipo de análisis es establecer prioridades en la utilización de los recursos asistenciales, de tal modo que se obtenga el mayor beneficio posible *para la comunidad*. Los análisis coste-eficacia, coste-efectividad centrado en el individuo, y coste-efectividad centrado en la comunidad pueden dar lugar a conclusiones distintas<sup>98</sup>.

### Conceptos relacionados con la efectividad clínica y la CER

Por lo tanto, la primera cuestión es cómo establecer la efectividad clínica de una intervención terapéutica. Consideremos el ejemplo citado antes del análisis de los beneficios terapéuticos de los bifosfonatos en la práctica asistencial, en relación con su capacidad para prevenir fracturas de cadera en personas ancianas<sup>101</sup>. Los paneles de expertos han concluido que el tratamiento con bifosfonatos es coste-efectivo (como se ha indicado antes, se trata de un término impreciso que

suele emplearse vulgarmente con la intención de identificar una intervención terapéutica como aquella que produce un beneficio determinado a un coste considerado adecuado para la sociedad; suele emplearse en términos absolutos explicativos, cuando en farmacoeconomía los análisis de coste-efectividad válidos suelen ser relativos, entendidos como coste-efectividad incremental). Pero analicemos el apartado de la efectividad. Las conclusiones sobre el tratamiento con bifosfonatos se basan en estimaciones de modelización de Markov, en análisis post-hoc de datos de los ensayos clínicos<sup>102</sup>. Pero el problema fundamental con los modelos de Markov es que los datos que sustentan la eficacia del tratamiento no indican su efectividad en práctica clínica<sup>101</sup>. En esencia, los modelos extienden las pruebas de la eficacia en ensayos clínicos como si fuesen aplicables en la práctica comunitaria. Con este supuesto, los expertos concluyen que, en lo que se refiere a la efectividad de los bifosfonatos, pueden generalizar el dato del 1-2% de reducción en el riesgo absoluto de prevención de fracturas, encontrado en el ensayo clínico, a una amplia población independientemente de la edad, sexo, comorbilidad, grado de osteopenia u osteoporosis, o historia previa de fractura en la comunidad. Esto está muy lejos de la realidad. La evidencia de que los bisfosfonatos pueden prevenir la fractura de cadera es muy limitada. Las reducciones significativas de fracturas de cadera se detectaron en una subpoblación restringida de mujeres (65 a 80 años con osteoporosis o fracturas previas), mientras que se carece de pruebas convincentes de la eficacia de los bifosfonatos entre los pacientes que más suelen sufrir de fracturas de cadera (personas de al menos 80 años de edad, y los que viven en hogares de ancianos)<sup>103</sup>. Y aunque la osteoporosis es considerada una enfermedad predominantemente femenina, aproximadamente el 40% de las fracturas relacionadas con la edad se producen en hombres de edad avanzada. A pesar de esto, todos los análisis actuales de la rentabilidad de bifosfonatos asumen una reducción universal en el riesgo de fractura entre todas las personas de más edad, equivalente a la encontrada en los ensayos clínicos. Pero si sólo se alcanzase una reducción del 32% en fractura de cadera con bisfosfonatos en los pacientes que cumplen criterios de inclusión de los ensayos clínicos, cuando se extrapolan los datos a la población general y se prescribe el tratamiento a todos los ciudadanos de edad ffl50 años sólo un 4,6% de las fracturas, y no un 32%, se evitarán por la administración de bisfosfonatos<sup>101</sup>.

Los estudios de efectividad tienen el complejo objetivo de evaluar los efectos beneficiosos de un tratamiento particular en el contexto de la práctica médica cotidiana. Aquí, en el mundo real, la población de pacientes es heterogénea, sus hábitos de vida y el cumplimiento terapéutico están menos controlados, y existen múltiples variables que pueden afectar a la evolución del paciente y al resultado final de la intervención terapéutica, aparte de la intervención en sí misma. El estudio de los efectos beneficiosos del tratamiento, por lo tanto, a menudo requiere muestras de población más grandes y en seguimiento monitorizado durante períodos de tiempo más largos. Además, el análisis de la efectividad requiere que los efectos de una serie de variables posibles y factores de confusión se analicen y consideren en lugar de eliminarse, como se hace al evaluar la eficacia. Esto incluye variables introducidas por los procesos de toma de decisiones complejas de cuidadores reales, los cuales pueden abarcar muchos factores que motivan decisiones respecto a la selección, inicio, el ajuste o el cese del tratamiento<sup>99</sup>.

Para algunos autores, tan prestigiosos como Rawlins, las pruebas que apoyen la efectividad clínica de una intervención terapéutica pueden ser de tipo experimental (es decir, ensayos clínicos controlados y aleatorizados), u observacional, mediante la recogida de información en condiciones asistenciales habituales<sup>104</sup>. Según este autor, en términos de efectividad lo importante no es el método en sí mismo, sino si el método utilizado es el adecuado para contestar a las preguntas planteadas<sup>39</sup>. Y, como se ha indicado antes, el uso exclusivo de datos basados en ensayos clínicos conlleva habitualmente el uso de métodos de jerarquización de la evidencia, defectuosos y potencialmente erróneos e inadecuados para contestar a preguntas prácticas concretas<sup>96</sup>. Se debe abrir por lo tanto una discusión sobre la metodología del trabajo en estudios de efectividad clínica, cuyo enfoque en CER se abordará más adelante.

La valoración de la efectividad clínica de una intervención terapéutica, para una población o comunidad, implica de modo crucial y esencial la evaluación de su efectividad *en comparación con las alternativas existentes*, incluyendo el tratamiento no activo como una alternativa. Por lo tanto, debe plantearse cómo identificar adecuadamente las alternativas con las que comparar el tratamiento en estudio. Esta decisión puede ser complicada desde el punto de vista de los responsables de la gestión asistencial, porque puede ser un error restringir los posibles comparadores sólo a aquellos que se han autorizado para tratar a los pacientes en los que se va a realizar el estudio de efectividad comparada<sup>104</sup>. Muchos tratamientos establecidos nunca se han autorizado para su empleo habitual, a pesar de que éste sea generalizado, como la dexametasona para el tratamiento del mieloma múltiple, o el uso fuera de indicación de medicamentos en pediatría: en la Unión Europea (UE), el cincuenta por ciento o más de los medicamentos utilizados en los niños nunca se han estudiado en esta población, sino sólo en los adultos, y no necesariamente en la misma indicación<sup>104,105</sup>. En la práctica, los informes de los expertos pueden orientar acerca de lo que constituye “el mejor tratamiento actual” con el fin de utilizarlo como el comparador adecuado. El consenso en este terreno no es fácil de alcanzar, y debiéramos estar alertas a los riesgos de una utilización sesgada.

**CER** Como se ha expuesto antes, la idea de que la decisión de utilizar un tratamiento u otro cuando existan dos o más opciones terapéuticas posibles debe basarse en los datos generados por la investigación científica, expuesta en las publicaciones médicas, no es nueva; sin embargo, en 2009 el Institute of Medicine (IOM) de USA afirmaba que más de la mitad de los tratamientos que recibían los pacientes eran administrados sin pruebas claras de eficacia clínica<sup>106</sup>. Por otra parte el número de opciones diagnósticas y terapéuticas está aumentando de modo exponencial en las últimas décadas. El ritmo de la innovación en la asistencia sanitaria nunca ha sido similar en la historia, y esta innovación está agregando constantemente nuevas opciones terapéuticas y elementos amplios y complejos en el análisis científico y económico de las intervenciones de atención de salud. La demostración de la persistente e inexplicable variabilidad terapéutica que se detecta en la práctica clínica, y las altas tasas de cuidado inapropiado percibido por el paciente, combinadas con el aumento de gastos, han alimentado una demanda cada vez mayor de la aportación de pruebas científicas de la efectividad clínica de las intervenciones y los sistemas de salud. No entender bajo qué circunstancias y para qué tipo de pacientes una intervención terapéutica es la

mejor, puede contribuir al aumento del coste de la atención asistencial, amenazar la seguridad del paciente, limitar su bienestar o mejora de la salud, o aumentar el riesgo de muerte<sup>100</sup>.

Así, para asegurar que la salud de la población mejora, en vez de disminuir, a medida que la información terapéutica asistencial se renueva y acumula, y se introducen nuevas opciones terapéuticas, deben generarse nuevas herramientas de análisis de efectividad comparada. Pero sólo recientemente el proceso de esta toma de decisiones basada en la literatura científica se ha constituido como método científico específico, así como es reciente la intención de hacer consistente el proceso en sí mismo, lo que ha conducido a que sobresalga y destaque en interés y relevancia en términos científicos y de gestión asistencial. En USA, el nuevo paradigma se conoce como CER<sup>99</sup>. CER se discute ahora diariamente a escala nacional (USA) e internacional por los médicos y los responsables políticos, y las iniciativas de financiación federal prácticamente han garantizado su interés en el ámbito de las ciencias de la salud, en el foro de políticas públicas, y en la vida de los pacientes en los próximos años. No obstante, mucha la literatura sobre CER ha sido escrita por y para epidemiólogos o responsables de las políticas, por lo que la mayoría de los médicos en ejercicio no son totalmente conscientes del significado práctico del término, de los principios fundamentales para su implementación, y la naturaleza y extensión del potencial impacto que puede suponer CER en su labor profesional y en la atención diaria del paciente. Salvo un cambio dramático en la opinión pública o el clima político, CER ocupará un papel de importancia progresiva en la práctica clínica y en la política de gestión de la salud en los próximos años, y por ello es fundamental que los médicos comprendan este nuevo paradigma de investigación<sup>100</sup>.

### Definición de CER

La definición de CER que se expone al inicio de esta monografía no es la única, incluso dentro de USA. Como se indicaba antes, el Federal Coordinating Council for Comparative Effectiveness Research define CER como la realización y la síntesis de investigación comparada de los beneficios y los daños de diferentes intervenciones y estrategias para prevenir, diagnosticar, tratar y monitorizar las condiciones sanitarias en condiciones de uso habitual<sup>1</sup>. El IOM es una organización independiente sin fines de lucro, establecida en 1970, que trabaja fuera del gobierno para proporcionar un asesoramiento imparcial y autorizado para los gestores de la atención sanitaria y el público. Para la IOM, la CER se define como la generación y síntesis de las pruebas comparativas de los beneficios y los daños de las distintas alternativas existentes para prevenir, diagnosticar, tratar y monitorizar una condición clínica o para mejorar la prestación de la asistencia<sup>106</sup>. El Colegio de Médicos Americano define CER como la evaluación de la efectividad clínica relativa, la seguridad y el coste de dos o más servicios médicos, productos sanitarios, tratamientos o procedimientos utilizados para tratar el mismo problema asistencial<sup>107</sup>.

Estas definiciones son similares pero no completamente equivalentes, y por ello es fundamental que los médicos observen las diferencias en cada definición. Estas diferencias pueden afectar la financiación, diseño, implementación y la presentación de informes en las investigaciones de CER patrocinadas por estas agencias.

### **Estructura y aspectos organizativos de CER en USA**

Con la aprobación del ARRA, el Congreso de USA asignó y comenzó a distribuir 1100 millones de dólares para acelerar los esfuerzos para desarrollar a gran escala el proyecto CER. Estos fondos, destinados a la puesta en marcha del programa a nivel nacional, se han dividido en tres agencias federales: 400 millones asignados a the Office of the Secretary of Health and Human Services, otros 400 millones a the National Institutes of Health (NIH), y 300 millones a the Agency for Healthcare Research and Quality (AHRQ)<sup>108</sup>. La estructura del este proyecto nacional CER incluye la creación de un centro CER dentro del AHRQ y el establecimiento de un Instituto de Investigación de Resultados Centrados en el Paciente (Patient-Centered Outcomes Research Institute; PCORI), una organización independiente sin fines de lucro para asesorar al gobierno federal con respecto a las prioridades de investigación. La legislación acompañante describe los poderes, deberes, métodos específicos, y la supervisión de las agencias creadas dentro del proyecto, y ofrece algunas restricciones con respecto al uso de CER realizada por estos organismos. Dada la gran cantidad de capital comprometido, los considerables esfuerzos legislativos, y la estructura administrativa integral erigida, se asume los médicos tendrán que entender el proyecto CER y aceptar que interferirá en su labor asistencial en los distintos sistemas de salud<sup>99</sup>.

Para encauzar los esfuerzos iniciales, se encargó a la IOM que desarrollase una lista de temas de investigación prioritarios a los que financiar de modo preferente. El Comité del IOM sobre la Priorización en la Investigación sobre Efectividad Comparada recolectó datos sobre temas prioritarios para CER, así como información sobre las poblaciones de estudio de interés, propuesta de comparadores, y metodología de estudio para cada tema. Se analizaron más de 20 000 solicitudes de los consumidores, asociaciones de pacientes, grupos de proveedores, aseguradoras, fabricantes, y académicos, con tres niveles de criterios de selección: criterios de equilibrio de la cartera de propuestas, criterios de indicadores de la efectividad y criterios destinados a priorizar los temas de trabajo<sup>106</sup>. La lista de prioridades incluye la investigación de enfermedades que afectan a gran parte de la población, así como las condiciones que tienen potencialmente graves efectos en subgrupos reducidos de pacientes, o los estudios en mujeres o en minorías étnicas.

Los temas de investigación incluyen la detección, prevención y vigilancia, como así como las intervenciones terapéuticas médicas, quirúrgicas y otras alternativas. La cartera final contiene una lista de los 100 principales temas, divididos por cuartiles, que recibirán fondos CER proporcionales. Entre las prioridades recomendadas en el informe de IOM, destacan las siguientes<sup>106, 109</sup>:

Sistemas de prestación de asistencia sanitaria: más que la investigación sobre qué servicios deben proporcionarse, la pregunta es cómo y dónde se deben proporcionar esos servicios de asistencia sanitaria.

- Disparidades raciales y étnicas en asistencia sanitaria.
- Enfermedades cardiovasculares y vasculares periféricas.
- Trastornos psiquiátricos.
- Enfermedades neurológicas.
- Cáncer.
- Salud de la mujer.

**Programas tipo CER  
en otros países.  
Diferencias con el  
programa CER**

Aunque el término CER es exclusivamente estadounidense, otros países han llevado a cabo experiencias de evaluación de tecnologías sanitarias o de intervenciones terapéuticas, que podrían ser próximos a la descripción y los fines de CER. Los ejemplos más notables de tales agencias incluyen el NICE para el Reino Unido, el Instituto Alemán de Calidad y Eficiencia en la Atención de la Salud, la Alta Autoridad para la Salud Francesa, la Agencia Canadiense de Medicamentos y Tecnologías en Salud, y el Departamento de Salud y Envejecimiento Australiano<sup>99,100,109</sup>.

En la mayoría de circunstancias, la investigación llevada a cabo por estos organismos ha surgido en respuesta a una demanda de orientación en las tomas de decisiones basadas en pruebas, y los resultados de estas investigaciones se han priorizado para cumplir con las necesidades de los responsables de la toma de decisiones asistenciales. Al igual que el sistema estadounidense propuesto, la mayoría de estos proyectos han recibido apoyo financiero de sus respectivos gobiernos. En general, todos estos sistemas comparten varios atributos principales, establecidos en su programa de trabajo, que han sido descritos en detalle por Chalkidou<sup>109</sup>: Propósito y objetivo declarados, selección de objetivos de la evaluación (intervenciones terapéuticas, tecnologías sanitarias, guías asistenciales....), establecimiento del procedimiento para establecer prioridades en la selección de temas, selección del tipo de investigaciones que se utilizarán para la evaluación de las pruebas, establecimiento de relaciones con estructuras de investigación académicas, establecimiento de relaciones con los sistemas de atención sanitaria, estructura presupuestaria y de financiación, consideración del coste y de la relación coste-efectividad en sus decisiones, la disposición normativa del estatus de las recomendaciones, y cómo se difunden, publican e implementan las conclusiones.

Entre las diferencias existentes entre los distintos sistemas indicados, y entre ellos y el programa CER, las más notables se refieren al estatus de las recomendaciones y el análisis e impacto sobre los costes y la financiación. Mientras que la mayoría de los sistemas no USA se encargan específicamente de la incorporación de datos de coste-efectividad en sus modelos de análisis y recomendaciones, los estadounidenses tienden a ver la idea de la consideración de los costes que pueden afectar la toma de decisiones médicas como algo negativo. Aunque los análisis de costes y de coste-efectividad o coste-utilidad no han sido excluidos de las recomendaciones del programa CER, se ve con recelo por algunos americanos el potencial para que las recomendaciones del programa patrocinado por el gobierno influyan en las políticas públicas con respecto a la cobertura de salud y de reembolso médico a los proveedores públicos y privados. Por consiguiente, los planes actuales en USA hacen hincapié en la capacidad del programa CER para informar sobre la actividad médica, pero quitando énfasis a los análisis de coste-efectividad<sup>99</sup>.

Las recomendaciones de la mayoría de programas fuera de USA se utilizan de un modo u otro por los respectivos gobiernos al establecer las políticas de salud y normativas de reembolso, pero esta práctica ha sido prohibida por el Congreso de Estados Unidos<sup>99</sup>.

Las diferentes entidades nacionales fuera de USA han adoptado diferentes modelos en relación con el estatus de las recomendaciones efectuadas. Por ejemplo, las guías NICE favorables a una tecnología asistencial son obligatorias y generan un

derecho constitucional legal para el paciente<sup>109,110</sup>. Esto significa que si un médico y su paciente deciden acceder a esa tecnología, deberán estar disponibles los fondos para dicho acceso, y asegurar el acceso libre sin coste para el paciente. No obstante, NICE no tiene autoridad para restringir el acceso a medicamentos, ni es responsable de la fijación de su precio<sup>110</sup>. En los países en los que las recomendaciones se utilizan para tomar decisiones sobre disponibilidades, precios y reembolsos, como en Francia o Australia, dichas recomendaciones pueden decidir si un medicamento está o no disponible para el sistema público sanitario<sup>109</sup>.

NICE es visto como un líder global en la incorporación de los valores sociales y el compromiso público en la elaboración de recomendaciones de política sanitaria. Sin embargo, la obtención de información sobre el valor que la población da a una intervención, utilizar esta información para apoyar la toma de decisiones, y evaluar el impacto de esas decisiones, se ha considerado, respectivamente, como un desafío, una dificultad y una imposibilidad<sup>110</sup>. En la actualidad, el proceso de toma de decisiones de NICE utiliza un límite definido de financiación, en función del AVAC obtenido con un medicamento concreto y de su coste, siendo el objetivo asegurar la mayor cantidad de beneficios de salud pública como sea posible dentro de un presupuesto específico determinado cada dos años. Otros valores sociales no se reflejan formalmente en este umbral, pero se espera que los comités de toma de decisiones de NICE lo tengan en cuenta en sus deliberaciones<sup>110</sup>. El análisis de coste-efectividad utilizando los AVAC como la principal medida de resultados de una intervención terapéutica (coste-utilidad) es también el método principal de evaluación en Canadá, Australia, Suecia y Francia. El umbral de financiación aceptable varía en los distintos países: en el Reino Unido es de 34.000 a 51.000 \$ por AVAC, mientras que en Suecia es de 77.000 a 85.000 \$ por AVAC (datos de 2014)<sup>111</sup>.

### **CER y la Unión Europea**

La Unión Europea (UE) ha promovido desde 1994 el establecimiento de prioridades, valoraciones de la efectividad de las intervenciones terapéuticas, y difusión de los resultados de las evaluaciones<sup>100</sup>. La UE estableció la Red Europea de Evaluación de Tecnologías Sanitarias (EUnetHTA) en 2006 para promover una mejor coordinación de los esfuerzos nacionales<sup>112</sup>. La Agencia Europea de Medicamentos (EMA) o los miembros de la red de agencias, autorizan la comercialización y la disponibilidad de los medicamentos evaluando los datos de beneficio y riesgo disponibles en cada caso. Como se indica en el Reglamento 726/2004 del Parlamento Europeo y del Consejo, “Los Estados miembros han desarrollado una evaluación de la eficacia relativa de los medicamentos con objeto de situar los nuevos medicamentos respecto a los ya existentes, en la misma clase terapéutica. De igual modo, en sus Conclusiones sobre medicamentos y salud pública, adoptadas el 29 de junio de 2000, el Consejo subrayó la importancia de la identificación de medicamentos con un destacado valor terapéutico añadido. *No obstante, no debe efectuarse este tipo de evaluación en el procedimiento de concesión de la autorización de comercialización, en el que deben primar los criterios fundamentales de eficacia*”<sup>113</sup>.

No obstante, ha habido una demanda creciente sobre la necesidad de que existan estudios comparativos directos entre los nuevos medicamentos propuestos para su comercialización y los medicamentos ya existentes, por lo general de un coste directo menor. Se proponía que, a pesar de las complejidades metodológicas que existan, las pruebas de eficacia comparada sean tenidas en cuenta formalmente en las decisiones de comercialización de un nuevo medicamento<sup>114</sup>. Así,

muchos encontraron motivos para considerar que se daba un valiente paso en esta dirección por la EMA al declarar que a partir de enero de 2014 publicaría los datos de los ensayos clínicos de los medicamentos cuya autorización estuviese valorando<sup>115</sup>. Después de que el proceso de autorización de comercialización se haya completado, los investigadores y otras personas que son independientes de los patrocinadores de los estudios, de los autores de publicaciones de informes, y de las agencias reguladoras, podrán revisar toda la información de todos los ensayos, incluyendo datos anonimizados de los pacientes, volver a analizar los datos, y llevar a cabo nuevos análisis. Para muchos investigadores, si la EMA cumple sus promesas, se abriría una nueva era en la que se consideren los datos de ensayos clínicos como un bien público<sup>116</sup>.

Sin embargo, para otros investigadores, esta medida, mereciendo su aplauso, presenta dos limitaciones: la primera, que no se aplica a los medicamentos autorizados antes de la fecha propuesta, ya que no se dispondría de datos para comparar los nuevos medicamentos con los ya existentes. Y la segunda, que, aún no proporcionará suficiente información para el análisis no sesgado de la efectividad comparada de los medicamentos<sup>117</sup>.

### **Los análisis de efectividad comparada en España**

En España, los datos de coste-efectividad y el beneficio clínico se han constituido como uno de los criterios básicos de financiación pública de medicamentos y fijación de precios, y su análisis es básico para la elaboración de los informes de posicionamiento terapéutico de los medicamentos.

La ley 29/2006, de garantías y uso racional de los medicamentos y productos sanitarios, establece en la modificación actualmente vigente lo siguiente<sup>118</sup>:

“Artículo 89. Procedimiento para la financiación pública. Para la financiación pública de los medicamentos y productos sanitarios será necesaria su inclusión en la prestación farmacéutica mediante la correspondiente resolución expresa de la unidad responsable del Ministerio de Sanidad, Servicios Sociales e Igualdad (...). La inclusión de medicamentos en la financiación del Sistema Nacional de Salud se posibilita mediante la financiación selectiva y no indiscriminada teniendo en cuenta criterios generales, objetivos y publicados y, concretamente, los siguientes: (...) c) Valor terapéutico y social del medicamento y beneficio clínico incremental del mismo *teniendo en cuenta su relación coste-efectividad*.” Y, más adelante, se establecen los mismos criterios para la inclusión de los productos sanitarios en la prestación farmacéutica.

“Artículo 89 bis. Criterios fundamentales de inclusión en la prestación farmacéutica (...) 2. Para la decisión de financiación de nuevos medicamentos, además del correspondiente análisis coste-efectividad y de impacto presupuestario, se tendrá en cuenta el componente de innovación, para avances terapéuticos indiscutibles (...)”

Pero en esta normativa no se hace referencia al proceso por el que se realizarán los análisis de coste-efectividad, ni el significado de este término. En diciembre de 2012 el Consejo Interterritorial del Sistema Nacional de Salud (SNS) aprobó la elaboración de “Informes de Posicionamiento Terapéutico”. El objetivo es obtener un único informe que sea asumido y reconocido por todo el Sistema Nacional de Salud, definiendo la financiación específica de nuevos medicamentos para los pacientes. En mayo de 2013 la Comisión Permanente de Farmacia del SNS aprobó una Propuesta de colaboración para la elaboración de los informes de posicionamiento terapéutico de los medicamentos<sup>119</sup>.

En esa propuesta de la Comisión se establece que en el proceso intervienen, al menos, las estructuras de evaluación de la Agencia Española de Medicamentos y Productos Sanitarios (AEMPS), de la Dirección General de Cartera Básica del Servicio Nacional de Salud y Farmacia, y las estructuras de las Comunidades Autónomas. Y que “parece deseable establecer un sistema que, con la participación de los actores señalados y mediante un sistema de evaluación en red basado en la evidencia científica, se comparta la realización de dichas evaluaciones permitiendo mantener el mismo beneficio final para cada uno de ellos, evitando la redundancia, manteniendo la coherencia en la evaluación y compartiendo los recursos de forma más eficiente -y por tanto sacando mayor provecho de ellos- elaborando un único informe de posicionamiento terapéutico que sea reconocible para todo el SNS”<sup>119</sup>. Y se especifica que “Los informes de posicionamiento terapéutico contendrán, en una primera fase, la *evaluación de la efectividad y seguridad comparada*, así como los criterios de uso y seguimiento”<sup>119</sup>. Pero en ninguna parte del texto se explica la metodología o los criterios que se seguirán para realizar estos análisis de efectividad comparada; solamente se indica que “El objetivo de los informes es ofrecer, más allá de la autorización del medicamento, información relevante, realista y no sesgada de la posición que el nuevo medicamento ocupa en el mercado en comparación con otros medicamentos o medidas de salud ya existentes” y que los informes para cada medicamento serán elaborados “por un Grupo de Evaluación que redactará el informe final que será presentado al Grupo de Coordinación y Posicionamiento Terapéutico, y que estará integrado por la AEMPS y dos Comunidades Autónomas más. Cada uno de los componentes del grupo de evaluación contará con el apoyo de profesionales que ellos mismos determinen con experiencia clínica en el área terapéutica objeto del informe”<sup>119</sup>.

Así pues, en España, los análisis de coste-efectividad tienen una repercusión directa en el proceso de su financiación pública, y, a través de los informes de Posicionamiento Terapéutico, establecer la financiación específica de nuevos medicamentos para los pacientes. La estructura dispuesta para la elaboración de estos informes está descrita de modo impreciso, y, entre las características de otros sistemas tipo CER consideradas por Chalkidou<sup>109</sup>, se pueden encontrar algunas (propósito y objetivo declarados, selección de objetivos de la evaluación, consideración del coste y de la relación coste-efectividad en sus decisiones, disposición normativa del estatus de las recomendaciones, cómo se difunden, publican e implementan las conclusiones, establecimiento de relaciones con los sistemas de atención sanitaria), pero no otras (procedimiento para establecer prioridades en la selección de temas, selección del tipo de investigaciones y procedimientos que se utilizarán para la evaluación de las pruebas, establecimiento de relaciones con estructuras de investigación académicas).

### **Las implicaciones de CER en la investigación clínica. La metodología de CER**

Una vez considerados todos los aspectos administrativos, organizativos, de gestión y financieros de los programas CER, es necesario volver a los orígenes y a los aspectos fundamentales. Se trata de que la decisión entre dos o más opciones terapéuticas posibles debe basarse en los datos generados por la investigación científica. Y es importante no equivocarse durante el proceso de la realización y la síntesis de investigación comparada de los beneficios y los daños de diferentes intervenciones y estrategias para prevenir, diagnosticar, tratar y monitorizar las condiciones sanitarias en condiciones de uso habitual. Parece que se exige la máxima perfección posible a las empresas innovadoras en el proceso de gene-

ración de los datos científicos de eficacia, pero no está tan claro si se exige una calidad óptima a los responsables de la gestión asistencial durante el proceso de análisis de esos datos y de los procesos por los que se realizan las estimaciones de efectividad. Se ha indicado antes cómo muchas guías terapéuticas pueden establecer recomendaciones terapéuticas de modo inadecuado, criticable desde el punto de vista científico, y cuya aplicación a la práctica asistencial puede no ser adecuada. ¿Cómo se puede garantizar que este nuevo tipo de recomendaciones será científicamente correcto y útil en la práctica asistencial? ¿Tendremos que aceptar los argumentos de autoridad, y no siempre de autoridad de carácter científico, sino en ocasiones meramente político?

Desde el punto de vista del proceso de investigación en CER, ya se ha señalado antes que el actual paradigma de la MBE se considera por algunos autores limitado para elaborar procesos de toma de decisiones clínicas. El paradigma de MBE, basado en el análisis de ensayos clínicos, y apoyado en metaanálisis y revisiones sistemáticas, se ha mostrado como una vía aceptable y eficiente para evaluar la eficacia de las intervenciones terapéuticas. Pero, precisamente por las características de los elementos en que se basa, es más difícil aceptar la MBE como un proceso adecuado para elaborar conclusiones sobre la efectividad de la intervención, y para la generalización de las conclusiones<sup>96,99,101,104</sup>. En el programa CER, el Colegio de Médicos Americano sugiere que pueden requerirse metodologías de investigación más allá de los tradicionales ensayos clínicos controlados, para que las conclusiones de la CER sean adecuadas<sup>120</sup>.

Se ha prestado sorprendentemente poca atención a lo que puede ser una pregunta crítica a los informes de CER: ¿Cómo podemos asegurar que sus resultados podrán mejorar significativamente la calidad y seguridad de la atención sanitaria recibida por el promedio de los pacientes? ¿No se necesitan también métodos basados en pruebas para descubrir y describir cómo los resultados de los ensayos clínicos y CER pueden ser implementados y se incorporen de modo eficiente a la práctica habitual?<sup>121</sup>. Para muchos autores hay que empezar analizando los desafíos metodológicos a los que se enfrenta la investigación de efectividad que evalúa el impacto de distintas intervenciones sobre la salud del paciente. IOM, en su informe sobre priorización, señala que la CER comprende un amplio espectro de métodos de investigación establecidos y emergentes, incluyendo ensayos clínicos, estudios observacionales, y las revisiones sistemáticas de las pruebas disponibles, y que las prioridades del programa CER deben balancearse con las principales metodologías utilizadas<sup>106</sup>.

Aunque el diseño de estudios válidos de efectividad puede ser más complejo que la determinación de eficacia en entornos de investigación altamente controlados, la esperanza es que CER, si se realiza correctamente, puede apoyar de un modo más correcto las decisiones asistenciales que mejoren la salud de los pacientes. Los datos de CER y deben ser válidos, no sesgados, relevantes, factibles y viables. Es absolutamente fundamental el principio de que las propuestas sobre decisiones médicas requieren pruebas adecuadas de efectividad, calidad y accesibilidad de las intervenciones asistenciales que se comparan. Para lograr estos objetivos, los enfoques actuales de la realización de la investigación clínica y de los servicios de salud pueden no ser suficientes<sup>100</sup>.

Los estudios experimentales deben seguir siendo la fuente fundamental de información CER, y para responder adecuadamente a las preguntas que requie-

ren estudios experimentales es de vital importancia desarrollar los diseños y la infraestructura que generará información creíble y relevante del modo más rápido y económico posible. En 2012 se ha dado un paso relevante en la valoración y la perspectiva de este camino a seguir, cuando metodólogos y otros investigadores fueron invitados a presentar innovaciones en métodos idóneos para la investigación CER; específicamente, métodos que pueden solventar algunas de las diferencias relevantes en los resultados de ensayos clínicos aleatorizados de eficacia y los estudios observacionales de efectividad realizados en la asistencia clínica habitual<sup>122</sup>.

### ***Ensayos clínicos aleatorizados y CER***

Sin duda alguna los ensayos clínicos aleatorizados son el método más riguroso y valioso diseñado para la investigación clínica, y han proporcionado los datos más precisos, fiables y útiles sobre la eficacia de las intervenciones terapéuticas. Los ensayos clínicos aleatorizados, al eliminar o reducir al mínimo los sesgos y las variables de confusión, proporcionarían los datos de mayor calidad en términos de validez interna. Por lo tanto, y como han sugerido distintos autores, no parece razonable abandonar el modelo de ensayo clínico aleatorizado en la investigación clínica, pero parece necesario que, para su aplicación a estudios CER, se introduzcan cambios en los modelos tradicionales de los ensayos clínicos aleatorizados. Tal y como se llevan a cabo actualmente, desde el punto de vista de la CER son ineficientes, complejos, costosos, excluyen subgrupos relevantes de pacientes en aras de la búsqueda de la homogeneidad de los grupos comparados, no suelen comportar comparaciones directas entre distintas opciones terapéuticas, las variables de medida de resultados son frecuentemente subrogadas y a corto plazo, y requieren mucho tiempo para su ejecución<sup>123</sup>. Así pues, tomar decisiones sobre efectividad basándose solamente en estos ensayos puede ser erróneo. Pongamos un ejemplo: en la fibrilación auricular, varios ensayos clínicos aleatorizados y de gran tamaño han comparado la eficacia de los nuevos anticoagulantes (apixaban, endoxaban, dabigatran, y rivaroxaban) frente a warfarina. En conjunto, los resultados reflejan el éxito de los nuevos anticoagulantes, ya que éstos tienden en general a reducir el riesgo tanto de la variable principal (ictus y embolismo sistémico) como del objetivo principal de seguridad (hemorragias graves)<sup>124</sup>. La siguiente pregunta lógica es si existen diferencias reales entre los distintos nuevos anticoagulantes en su eficacia y seguridad, y, si es así, ¿qué nuevo anticoagulante parece ser mejor? Cuando se realizan comparaciones indirectas, utilizando las diferencias con warfarina como referencia común en los distintos ensayos, parece que las dosis altas de dabigatran proporcionan la mayor reducción en el riesgo de ictus y embolismo sistémico, mientras que apexaban tiene una eficacia similar y produce menos hemorragias graves<sup>124</sup>. Pero estas afirmaciones no son fiables porque se basan en comparaciones indirectas, valorando las diferencias de cada nuevo anticoagulante frente a warfarina en el ensayo clínico individualmente realizado. Esta comparación indirecta se lleva a cabo sin considerar que los ensayos de los nuevos anticoagulantes se diferencian entre sí en los criterios de inclusión de los pacientes, en las definiciones de los objetivos, en el enmascaramiento de los estudios, o en el grado de control de la coagulación por warfarina ajustada según la International Normalized Ratio (INR), pero no está claro en qué medida éstas diferencias afectan a la validez de las comparaciones indirectas<sup>124</sup>.

Así pues, CER no debe sustituir el paradigma tradicional de investigación, sino complementar el paradigma tradicional con modelos de investigación que proporcionen información útil y válida para la toma de decisiones post-reguladoras de una forma sistemática y metodológicamente más robusta.

**Cambios en el  
modelo de estudio:  
los ensayos clínicos  
aleatorizados  
pragmáticos (ECAP)**

Por lo tanto, para cumplir las funciones básicas de CER, se han de introducir características más pragmáticas de diseño de los ensayos clínicos. La actitud pragmática en el diseño de ensayos clínicos controlados fue descrita inicialmente en 1967 (y publicada de nuevo en 2009) por Schwartz y Lellouch durante el desarrollo de los debates sobre el papel del placebo y sobre el diseño más adecuado de los ensayos para producir cambios en la práctica clínica<sup>125</sup>. En su artículo, introdujeron la distinción entre las actitudes “pragmáticas” y las “explicativas” como intenciones alternativas subyacentes al diseño de cualquier ensayo clínico. El modelo de actitud pragmática, en el campo de los ensayos clínicos, son los ECAP.

Las pruebas experimentales controladas se realizan mediante ensayos explicativos, que responden a la pregunta de si el tratamiento puede ser útil, es decir, si puede demostrar su beneficio (eficacia) en circunstancias experimentales controladas<sup>126</sup>. A diferencia de los ensayos explicativos *el objetivo de los ECAP es medir la efectividad*, es decir, el beneficio de la intervención terapéutica en la práctica asistencial, *comparando directamente* distintas intervenciones terapéuticas utilizadas en una situación asistencial concreta<sup>126,127</sup>. El propósito habitual de un ECAP es ayudar a los médicos a decidir si es mejor una nueva intervención que otra de referencia (por ejemplo el tratamiento aceptado como el mejor en el momento actual), y también responder a las preguntas y satisfacer las necesidades de los gestores de recursos asistenciales en los procesos de la toma de decisiones post-comercialización<sup>99,126</sup>. Los ECAP pretenden proporcionar pruebas sobre si una intervención es efectiva, y si es o no más efectiva que otra, en el mundo real. En la práctica médica habitual los pacientes a menudo tienen comorbilidades, buscan atención médica en centros asistenciales que no siempre son académicos o de excelencia, y toman decisiones, junto con su médico, respecto a cuál es la intervención terapéutica más adecuada a su situación, de entre las opciones disponibles. En otras palabras, los ECAP solventan la principal limitación de los ensayos clínicos aleatorizados expuesta previamente: responden a la pregunta de si el tratamiento es útil en el entorno clínico asistencial habitual (efectividad), más allá de lo que indiquen las necesarias pruebas obtenidas en condiciones experimentales controladas.

Los ECAP se desarrollan en el entorno de la práctica clínica habitual, y los participantes pueden estar recibiendo al mismo tiempo otros cuidados asistenciales. De modo más general, los ensayos pragmáticos están diseñados para que los participantes incorporen la variabilidad que se observa en la práctica asistencial en la que se aplican los tratamientos autorizados, y esos participantes representen adecuadamente a los pacientes que recibirán dichos tratamientos en la práctica médica habitual. Por ese motivo, normalmente en los ECAP hay pocos criterios de selección distintos de la presencia de la enfermedad o el diagnóstico clínico de interés, y los participantes estarán en un grupo heterogéneo con respecto a sus características personales individuales. Una vez asignado a un grupo de tratamiento, el protocolo terapéutico puede ser adaptado a las necesidades del paciente, exactamente tal y como se aplicarán esos tratamientos en la práctica

asistencial<sup>126,127</sup>. Varias características que diferencian a menudo a los ECAP de los ensayos explicativos son, en los ECAP: la inclusión de comparadores activos, seleccionando alternativas clínicas relevantes como referentes/controles; que las medidas de resultados están centradas en el paciente, y que incorporan una gama amplia de pacientes (con diferentes co-morbididades, y diversos centros, condiciones asistenciales y características demográficas, todos ellos elementos implicados en la variabilidad terapéutica, cuyo control es una de las razones de ser del programa de CER)<sup>99,123,126</sup>.

Del mismo modo que sucede en los ensayos explicativos, los ECAP asignan aleatoriamente a los participantes al grupo de tratamiento, con el fin de garantizar un nivel suficiente en la validez interna del estudio. Pero, a diferencia de los ensayos explicativos, los ECAP presentan una elevada validez externa, una gran fiabilidad en la extrapolación de los resultados a la práctica asistencial<sup>125,126</sup>. El aspecto científico y práctico más atractivo de los ECAP es que su objetivo es lograr un equilibrio entre la validez interna, la relevancia y la pertinencia, manteniendo los beneficios de la aleatorización a la vez que incluye elementos de diseño que mejoran la validez externa<sup>128</sup>.

El creciente interés en los enfoques pragmáticos para el diseño de los ensayos clínicos se detecta por el aumento de la frecuencia de la publicación de artículos sobre los temas relacionados con ensayos pragmáticos, ensayos prácticos, pruebas en el mundo real, los ensayos de gestión y los ensayos de efectividad<sup>128</sup>. Se ha llevado a cabo una búsqueda en MEDLINE de los ECAP para ensayos que incluían el término “práctico” o “pragmático” en el título o el resumen, y el resultado muestra que, a lo largo las dos últimas décadas, el número de ECAP ha ido en aumento, de 43 en 1990 a 252 en 2010; sin embargo, la proporción de ensayos designados como “pragmático” es aún muy pequeña en relación con el total de ensayos clínicos<sup>128</sup>.

Dado el creciente, aunque incipiente, interés en los ECAP, ha sido necesario también definir el marco conceptual, la terminología y las normas de calidad para las publicaciones de ECAP. Durante los años 90 ya tuvo lugar un proceso de desarrollo de normas de calidad en la publicación de los ensayos clínicos en general. Se estableció que para entender y evaluar adecuadamente los resultados de un ensayo controlado aleatorio, los lectores deben comprender el diseño, la realización, el análisis y la interpretación de los datos del ensayo. Investigadores y editores presentaron la declaración CONSORT (Consolidated Standards of Reporting Trials) para ayudar a los autores y revisores a mejorar la publicación de los ensayos mediante dos procedimientos: el uso de un listado para realizar chequeos de evaluación del diseño del estudio, y un diagrama de flujo de los pacientes incluidos en el estudio. La declaración CONSORT se ha actualizado en 2001<sup>129</sup>. La lista actualizada incluye 22 elementos, seleccionados en base a que la evidencia empírica indicaba que la falta de información referida a ellos se asociaba con estimaciones sesgadas del efecto del tratamiento, o de la fiabilidad o la relevancia de los hallazgos. El diagrama de flujo de pacientes revisado representa información de 4 etapas de un ensayo (inclusión, asignación de la intervención, seguimiento y análisis).

Más recientemente se ha dado otro paso relevante, actualizando la *declaración CONSORT específica para los ECAP*<sup>130</sup>. Los autores de esta declaración recomiendan, para la presentación de los resultados de ensayos pragmáticos, ampliar ocho de

los puntos de la lista CONSORT: los antecedentes, los participantes, las intervenciones del estudio, los resultados, el tamaño de la muestra, el enmascaramiento, el diagrama de flujo de los participantes, y la validez externa del estudio (posibilidad de extrapolación y generalización de los resultados). Los autores proponen que, si se utilizan estos criterios, se facilitará a los lectores y gestores la evaluación de la validez de los resultados, el grado en que pueden ser aplicables a las distintas situaciones asistenciales que interesen para su aplicación, y la viabilidad de las intervenciones puestas a prueba en el ECAP<sup>130</sup>.

También se han desarrollado herramientas con el propósito de ayudar a los investigadores en el diseño de los ensayos, y facilitar el diálogo, ya en las fases iniciales del proceso de diseño de un ECAP, entre los investigadores y los que finalmente tomarán decisiones sobre la utilización de las intervenciones terapéuticas<sup>128</sup>. PRECIS (Pragmatic–Explanatory Continuum Indicator Summary) es una herramienta que constituye un intento de identificar y cuantificar las características del ensayo que diferencian los ensayos con diseño pragmático de aquellos con diseño explicativo<sup>131</sup>. PRECIS muestra de modo gráfico el grado de “pragmatismo” del estudio en 10 elementos estructurales definidos del diseño de los ensayos. La herramienta PRECIS hace visibles las opciones y equilibrios que los diseñadores del estudio decidan entre los 10 dominios del continuo explicativo-pragmático, cuando intentan balancear la relevancia en el mundo real, la viabilidad y la validez del estudio<sup>128,131</sup>.

***Cambios en el  
procedimiento de  
análisis de los datos:  
el papel potencial  
de la aproximación  
bayesiana adaptativa***

La estadística que estamos acostumbrados a utilizar es la estadística frecuentista, que es la que se desarrolla a partir de los conceptos de probabilidad y que se centra en el cálculo de probabilidades y los contrastes de hipótesis. La estadística frecuentista tiene como objetivo establecer conclusiones en base a la significación estadística y la aceptación o el rechazo de hipótesis, siempre dentro del marco del estudio que se esté realizando. En el análisis estadístico que pretende comparar la eficacia de un nuevo tratamiento frente a otro conocido, se utiliza únicamente la información obtenida en el ensayo. No existen subjetividades referentes a parámetros, puesto que se han fijado los criterios de decisión a priori y estos permanecen estáticos durante todo el estudio.

Como enfoque alternativo a la estadística frecuentista, aparece cada vez más en escena la estadística bayesiana, que se diferencia de la estadística frecuentista básicamente en la incorporación de información externa al estudio que se esté realizando, de manera que, si tenemos una estimación de la probabilidad de que ocurra un suceso, su valor será modificado cuando dispongamos de esa información<sup>132</sup>. Así pues, las fuentes de información “a priori” se ven transformadas en probabilidad “a posteriori” cuando se dispone de los resultados del estudio, y se utilizan a continuación estos datos para realizar la inferencia. La escuela frecuentista tradicional ha proporcionado una sólida base para la estadística médica. Pero aunque la división artificial de los resultados en “significativo” y “no significativo” puede ser adecuada para decisiones dicotómicas, como la autorización o no de un medicamento, no parece que sea el mejor modelo para comparar intervenciones a medida que las pruebas y los datos se acumulan con el tiempo, como ocurre en un sistema de atención médica dinámico<sup>123,132</sup>. El mundo real de la CER puede presentar más ruido, más variabilidad de referencia que un ensayo clínico estándar. Se requeriría por lo tanto en estos ensayos, si se utilizase la esta-

dística frecuentista, un tamaño muestral muy elevado, en parte por el aumento adicional de la variabilidad, y en parte porque se comparan varios tratamientos activos cuya eficacia puede diferir en magnitudes relativamente pequeñas<sup>123</sup>.

Un diseño que utilice características que cambian o se adaptan en respuesta a la información generada durante el ensayo puede ser más eficiente que los enfoques estándar. El enfoque bayesiano, usando enunciados probabilísticos formales de incertidumbre, puede combinar todas las fuentes de información internas y externas al estudio, mientras controla la probabilidad de falsos positivos y falsos negativos en las conclusiones. Los enfoques analíticos bayesianos y adaptativos pueden reducir el tamaño de la muestra, el tiempo y el coste requeridos para obtener información relevante mediante la incorporación de evidencia externa de alta calidad en el diseño de la CER. La experiencia con los enfoques bayesianos adaptativos ha ido creciendo en los últimos años. Los diseños bayesianos son habituales en el diseño de ensayos de cáncer en fase inicial, en 2005 la FDA dio a conocer un documento guía de orientación para el uso de métodos bayesianos en los ensayos de productos sanitarios, y el Center for Drug Evaluation and Research and Center for Biologics Evaluation and Research de la FDA ha aceptado los diseños bayesianos de ensayos clínicos<sup>133</sup>. Parece que es el momento adecuado para el uso de diseños adaptativos bayesianos en ensayos clínicos de efectividad comparada<sup>134-136</sup>.

### **Estudios observacionales y CER**

El paradigma CER reconoce también el valor potencial de los estudios observacionales en la evaluación global de la efectividad de las distintas alternativas terapéuticas. Para ello es necesario controlar los sesgos potenciales de estos estudios, su objetividad, las variables de medida, su diseño y la logística intrínseca de cada modelo. Si bien estos aspectos generalmente se conocen bien y se han discutido para la evaluación de los ensayos clínicos aleatorizados, se ha prestado menos atención a sus implicaciones para la utilización de los estudios observacionales en el contexto de la CER<sup>99</sup>.

Los análisis de CER han conllevado un renovado interés por el papel de los estudios observacionales en el proceso de evaluación de los beneficios y los riesgos de las distintas intervenciones terapéuticas. Los estudios observacionales comparan los resultados entre los pacientes que reciben diferentes intervenciones, donde la decisión de la intervención terapéutica que recibe un paciente concreto no ha sido el fruto de un proceso de aleatorización, sino de la decisión médica que ha considerado que ése era el tratamiento adecuado para ése paciente individual. Por lo general el proceso de asignación a distintos tratamientos está determinado por la variación natural en el proceso rutinario de la toma de decisiones en la atención clínica habitual; pero los estudios observacionales también pueden hacer uso ventajoso de modificaciones en la atención por decisiones a otro nivel, como cambios en la política asistencial, cambios en la estructura de la unidad asistencial o cambios en las relaciones con los proveedores, que conducen a variaciones en la exposición a distintas intervenciones terapéuticas entre los grupos de pacientes<sup>137</sup>.

En los estudios observacionales pueden incluirse los pacientes en función de la exposición (por ejemplo, el tipo de intervención) utilizando un diseño de cohortes, o en función del resultado mediante un diseño de casos y controles<sup>34</sup>. Los estudios de cohortes pueden ser realizados reclutando a los pacientes de forma

prospectiva, de tal modo que los participantes se incluyen en el estudio a partir del momento en que tiene lugar la exposición al tratamiento, o de forma retrospectiva, cuando se produjo la exposición antes de que se identificase a los potenciales participantes. En España, la normativa que regula estos estudios está vigente desde 2009<sup>138</sup>.

Las fortalezas y debilidades de los estudios observacionales para investigación clínica de la efectividad han sido objeto de debate durante décadas<sup>124,137,139</sup>. Debido a que el coste incremental de incluir un participante adicional en el estudio es generalmente bajo, los estudios observacionales a menudo tienen un número relativamente elevado de participantes que pueden ser más representativos de la población general. Si la población del estudio es amplia, los resultados pueden ser más generalizables y aplicables a la práctica asistencial en el mundo real, y puede permitir valorar las diferencias de su efectividad en distintos subgrupos de pacientes. Esta ventaja es particularmente importante para una de las justificaciones del proyecto CER, comprender la eficacia del tratamiento en las poblaciones vulnerables, tales como las minorías raciales o en otros grupos frecuentemente infrarrepresentados en los ensayos clínicos aleatorizados<sup>137</sup>.

Sin embargo, los estudios observacionales también tienen limitaciones importantes<sup>124,137</sup>. Los estudios observacionales sólo son factibles si la intervención en evaluación ya está siendo utilizada en la práctica clínica, por lo que no son adecuados para la evaluación de nuevos fármacos o nuevos productos sanitarios, hasta que su empleo asistencial sea habitual. Otra limitación importante en los estudios observacionales es el riesgo de sesgos, que incluyen los sesgos de realización, de observación, de memoria y de selección<sup>34,137</sup>. El sesgo de realización se produce cuando la prescripción de un tipo de intervención se asocia con niveles de rendimiento generalmente mejores de la unidad de salud en la que se realiza el estudio (es decir, la calidad de la atención de la salud es mayor), lo que hace difícil concluir si los mejores resultados en salud que concluya el estudio son atribuibles a la intervención en sí o a la buena calidad de la asistencia sanitaria prestada en ese centro concreto. El sesgo de observación ocurre cuando los observadores valoran de forma diferente a los distintos grupos, o lo hacen los distintos observadores participantes, que pueden diferir además en la prestación asistencial que realizan a sus distintos grupos de pacientes. Una preocupación particular de la CER en relación con nuevos tratamientos es que algunas intervenciones nuevas pueden ser más susceptibles de ser utilizadas en pacientes en los que los tratamientos habituales considerados de primera elección han fracasado; estos pacientes probablemente responderán peor a cualquier terapia, y en consecuencia los resultados de la intervención en evaluación, debido al tipo de pacientes en los que se ha utilizado, serán peores<sup>137</sup>.

La falta de homogeneidad entre los grupos focaliza la preocupación más importante para todos los investigadores en el análisis de la validez interna de los estudios observacionales, y se produce cuando los grupos de intervención difieren en características que se asocian con el resultado de interés. Estas diferencias pueden ocurrir debido a que esa característica que forma parte de la decisión sobre qué tratamiento recomendar (se pueden recomendar distintos tratamientos en función de la gravedad de la enfermedad, de la presencia o ausencia de otras enfermedades, de alteraciones funcionales o de otros tratamientos, de la edad, etc.). El problema es que, en ausencia de asignación aleatoria, existe siempre un riesgo potencialmente elevado de que exista falta de homogeneidad<sup>138-141</sup>. Aun-

que existen puntuaciones de tendencias o análisis de covarianza que pueden ajustar por factores de confusión, sería ingenuo suponer que el potencial riesgo no existe, o se resuelve mediante un análisis estadístico. Disponer de un elevado número de pacientes podría limitar el sesgo, pero no se puede esperar que en las bases de datos utilizadas se hayan registrado todas las razones complejas que puedan explicar por qué un paciente recibe un tratamiento y otro paciente recibe el tratamiento alternativo.

Un ejemplo de la vulnerabilidad a los sesgos y variables de confusión de los estudios observacionales es el de los estudios observacionales epidemiológicos tanto prospectivos como retrospectivos que aportaban pruebas abundantes que demostraban que la terapia con estrógenos reducía en un 35-50% el riesgo de enfermedad cardiovascular primaria en mujeres posmenopáusicas<sup>142</sup>. Sin embargo, un ensayo clínico aleatorizado del grupo Women's Health Initiative, cuyo objetivo primario fue evaluar la incidencia de enfermedad coronaria en mujeres postmenopáusicas histerectomizadas tratadas con estrógenos, demostró que dicha reducción no era cierta<sup>143</sup>.

En 1997 tuvo lugar una polémica internacional que dejó muy mermada la credibilidad de los estudios observacionales para la sustentar per se la toma de decisiones en la práctica asistencial. Un subcomité creado específicamente con ese fin del Liaison Committee of the World Health Organisation and the International Society of Hypertension emitió un dictamen formal y final en respuesta a la acumulación de publicaciones generada en los años anteriores, que indicaban que el uso de fármacos calcio-antagonistas en el tratamiento de la hipertensión arterial se asociaba a un aumento de riesgo de patología cardiovascular coronaria<sup>144</sup>. En su respuesta a esta polémica, el subcomité rechazaba la validez de los datos de riesgo presentados, con la siguiente argumentación: en primer lugar, el subcomité analizó los datos de los estudios observacionales causantes de la polémica; respecto a los datos de los estudios de casos y controles, se consideraba que, desde el punto de vista metodológico, la principal preocupación del subcomité era el elevado riesgo de errores sistemáticos que presentan en general este tipo de estudios y que pueden afectar a los resultados obtenidos. Para el subcomité era posible, incluso probable, que la decisión de prescribir un antagonista del calcio específicamente para el tratamiento de la hipertensión arterial se hubiese tomado precisamente porque los pacientes presentaran factores de riesgo de cardiopatía coronaria, lo cual hubiese inducido a su prescripción (los calcio-antagonistas están indicados en el tratamiento de las coronariopatías). Así, estos estudios no evitarían el riesgo de estar comparando distintas intervenciones en grupos de población no homogéneos. Además, para el subcomité, aunque los análisis estadísticos de estos estudios incluyeron ajustes para algunos factores de confusión potencial, eso no garantizaba que la falta de homogeneidad no fuese una explicación alternativa plausible de los hallazgos observados<sup>144</sup>. También se planteaba la misma limitación para desestimar la credibilidad del aumento de riesgo de cardiopatía coronaria detectada en varios estudios de cohortes. En segundo lugar, el subcomité evaluó los ensayos clínicos aleatorizados realizados con calcio-antagonistas en el tratamiento de la hipertensión arterial; ninguno de los ensayos se diseñó para evaluar específicamente como objetivo primario el efecto de estos medicamentos sobre la patología cardiovascular coronaria; ni individualmente ni en conjunto, los ensayos indicaban una plausible diferencia en el riesgo de patología cardiovascular coronaria entre los calcio-antagonistas

y el placebo u otros tratamientos; estos mismos juicios se aplicaban también al subgrupo de pacientes con enfermedad vascular coronaria previa<sup>144</sup>. Caso cerrado.

Así pues, tampoco los meta-análisis de los estudios observacionales solucionan sus debilidades; por el contrario, los meta-análisis arrastran, en un proceso en el que se pierde la trazabilidad, los errores, sesgos y limitaciones de cada estudio particular. Y las consecuencias de ello son impredecibles.

Se pretenden implementar distintos tipos de análisis estadístico que solvente los sesgos de los estudios observacionales, además de la utilización de métodos clásicos como los ajustes multivariantes mediante modelos de regresión<sup>137</sup>. Un tipo de aproximación que ha suscitado gran interés en este sentido es el análisis o *puntuación de propensiones* (Propensity score matching), técnica estadística de coincidencia que intenta estimar el efecto de covariables predictoras de la probabilidad de recibir un tratamiento concreto. El uso de la técnica de puntuación de propensiones ha incrementado su popularidad en la investigación de efectividad en oncología en la última década; algunas revisiones indican que sólo en 2011 se han publicado en el Journal of Clinical Oncology seis artículos que la utilizan<sup>137</sup>.

Otro tipo de análisis, proveniente del campo de la economía y cuya aplicación al campo de los estudios observacionales en salud también se incrementa, es el método de *variable instrumental*, utilizado para estimar relaciones causales cuando no se dispone de experimentos controlados, o éstos no son factibles. En la práctica clínica pueden existir múltiples variables conocidas o desconocidas que influyen en la relación entre el factor evaluado y el desenlace. Las variables instrumentales son uno de los métodos desarrollados para controlar el efecto de los factores de confusión, medidos o no, simulando el efecto del azar en la asignación aleatoria que se realiza en los ensayos clínicos<sup>137,145</sup>.

Ante el debate acerca de la calidad de los estudios observacionales de efectividad, y su utilidad para aportar información útil y fiable a la CER, se ha planteado elaborar sistemas de evaluación de dicha calidad, además de exigir a estos estudios que en su publicación cumplan con los criterios de calidad de consenso para cualquier estudio observacional enumerados en la declaración STROBE (Strengthening the Reporting of Observational Studies in Epidemiology)<sup>146</sup>. Una de estas listas diseñadas para chequear la calidad de los estudios observacionales de efectividad comparada es la lista GRACE<sup>147</sup>. En el artículo que desarrolla y valida el método GRACE, los autores concluyen que este método puede ayudar a descartar estudios observacionales de baja calidad en el análisis de la CER, y que en conjunto este método discrimina mejor que las opiniones de expertos, tanto individuales como consensuadas<sup>147</sup>.

### Registros de pacientes

Finalmente, los estudios observacionales están absolutamente condicionados por el proceso de recogida de datos, la calidad y amplitud de esos datos, y su validez en las bases de datos utilizadas. La AHRQ ha publicado en poco tiempo la 3ª edición de un manual (la primera edición fue en 2007), cuyo objetivo es servir de guía para el diseño y uso de los registros de pacientes con fines de política científica, investigación clínica y cuidados de salud<sup>148</sup>. Esa Guía del Usuario está destinada a apoyar el diseño, la implementación, el análisis, la interpretación y la evaluación de la calidad de los registros creados para mejorar la comprensión de los resultados de las intervenciones terapéuticas en el paciente. En la Guía definen un registro de pacientes no como se conoce clásicamente para los estu-

dios epidemiológicos, sino como un sistema organizado que utiliza métodos de estudios observacionales para la recopilación de datos uniformes para evaluar los resultados específicos para una población definida por una enfermedad particular, condición, o una exposición, y que sirve a propósitos preestablecidos científicos, clínicos, o de política asistencial. La guía se centra en los registros creados para uno o más de los siguientes fines: describir la historia natural de la enfermedad, para *determinar la efectividad clínica* o la relación coste-efectividad de las intervenciones en salud, medir o monitorizar la seguridad y el riesgo, y/o cuantificar la calidad de la asistencia prestada. Y entienden que si se diseñan y ejecutan correctamente, los registros de pacientes pueden proporcionar una adecuada visión del mundo real de la práctica clínica, de los resultados que la asistencia produce en los pacientes, de su seguridad y *de la efectividad comparada*<sup>148</sup>. Realmente se trata del primer documento sistematizado que aborda la creación de registro de pacientes y de datos útiles para los estudios observacionales dirigidos a la CER, y lo hace de modo exhaustivo desde los ámbitos de la legalidad, la ética y el método científico.

### ***El cuidado centrado en el paciente en la CER***

La efectividad de las intervenciones debe basarse en la consecución de objetivos relevantes de salud para el paciente, cuyo máximo exponente, como se ha indicado previamente, son los AVAC, que incluyen una valoración subjetiva del paciente sobre su estado de salud. Hasta ahora, la participación de los pacientes en la evaluación de los objetivos y resultados de las intervenciones terapéuticas y la asistencia sanitaria ha sido escasa, pero está adquiriendo progresivamente más relevancia. La “*medida de resultados basada en el paciente*” (patient-based outcome measure) es un término que hace referencia a una serie de cuestionarios, programas de entrevistas y otros métodos relacionados con la evaluación de la salud, la enfermedad y los beneficios de las intervenciones de atención de la salud desde la perspectiva del paciente. La medida de resultados basada en el paciente, abordando constructos tales como la calidad de vida relacionada con la salud, el estado de salud subjetivo, o el estado funcional, se utilizan cada vez más como objetivos primarios o secundarios en los ensayos clínicos<sup>149</sup>.

Este término no se debe confundir con el de “medida de resultados notificada por el paciente” (PROM), que se refiere a los medios de recopilar información sobre la eficacia de la atención prestada a los pacientes, según la percepción de los propios pacientes, y que ha sido promovida por agencias como la FDA o el Instituto Nacional de Salud británico<sup>150,151</sup>. Al abrigo de estas recomendaciones se han desarrollado, en muchas ocasiones gracias a la financiación de la industria farmacéutica y no siempre de manera ordenada, numerosos instrumentos que persiguen evaluar las mejoras en síntomas o la calidad de vida de una determinada intervención en el paciente<sup>152</sup>.

Tras la creación del PCORI, una de las tareas que se encomendaron a este instituto fue ayudar a los pacientes, a los médicos y a los gestores políticos en la toma de decisiones de salud, mejorando la calidad y la pertinencia de los datos científicos que indican el modo en que sus enfermedades o trastornos de salud puedan prevenirse, diagnosticarse, tratarse y controlarse de modo eficaz y adecuado. La ayuda del PCORI debía realizarse mediante la investigación y la síntesis de las pruebas disponibles, basadas en estudios adecuados, considerando la variabilidad existente en subpoblaciones de pacientes, y mediante la difusión de las conclu-

siones de la CER con respecto a los resultados de salud, la efectividad clínica, y la adecuación de los tratamientos médicos<sup>111</sup>.

PCORI justifica su actividad indicando que los pacientes, sus familias y los médicos se enfrentan a una amplia gama de opciones complejas y a menudo confusas cuando se trata de hacer frente a sus problemas de salud y de asistencia sanitaria. Se necesita información fiable para decidir qué opción es la mejor para ellos, pero, desafortunadamente, la investigación médica tradicional, pese a todos los notables avances en la atención que produce, no ha sido capaz de responder a muchas de las preguntas a las que los pacientes y sus médicos se enfrentan a diario. Y aun cuando la información fiable existe, no es siempre accesible<sup>153</sup>.

PCORI indica que se creó para ayudar a abordar estos desafíos. Para los pacientes, esta estrategia significa que debe proporcionar información acerca de cuáles serían las intervenciones más beneficiosas para ellos, teniendo en cuenta sus preocupaciones particulares, circunstancias y preferencias<sup>153</sup>.

PCORI define y establece además la PCOR a través de su Comité de Metodología, con aportaciones de los interesados. La definición comienza con la finalidad última de la investigación, estableciendo que PCOR ayuda a las personas y sus cuidadores a comunicar y realizar informes sobre decisiones de cuidados de salud, lo que permite que sus intereses sean escuchados en la evaluación de las opciones asistenciales. Algunos ejemplos de preguntas de los pacientes que deberían contestarse son<sup>111</sup>:

- Dadas mis características, condiciones y preferencias personales, ¿qué puedo esperar que me suceda?
- ¿Cuáles son mis opciones y cuáles son los beneficios y riesgos potenciales de esas opciones?
- ¿Qué puedo hacer para mejorar los resultados que son más importantes para mí?

Se han realizado varios esfuerzos para identificar las prioridades de PCOR en USA. Antes se han comentado las prioridades de investigación establecidas por el IOM. Más que identificar las prioridades o exigir a PCORI que desarrolle una agenda de investigación basada en las prioridades de la OIM, PCORI ha establecido cinco áreas prioritarias de investigación para alcanzar sus objetivos de cambiar la forma en que la investigación se realiza y se comunica. Las cinco áreas prioritarias son<sup>111</sup>: la evaluación de la prevención, diagnóstico, y opciones de tratamiento; la mejora de los sistemas de salud; la evaluación de las disparidades en las prácticas clínicas; la comunicación y la difusión de la investigación; y, finalmente, el impulso de PCOR y de la metodología de la investigación en la CER



## A modo de epílogo

---

Aunque los avances en la ciencia médica han dado lugar a mejoras continuas en los resultados del cuidado de salud y médicos, las pruebas de la efectividad comparada de opciones terapéuticas alternativas siguen siendo insuficientes para responder a las preguntas que se plantean en la atención médica que se presta en el mundo real y la toma de decisiones políticas relacionadas con la salud. La consecuencia es una asistencia médica con frecuencia inferior a la mejor posible, no tan eficiente como debiera, y cuyos costes pueden llegar a ser difícilmente sostenibles<sup>123</sup>. Existe una absoluta y perentoria necesidad de disponer de información y de pruebas fiables y adecuadas que permitan tomar decisiones terapéuticas teniendo en cuenta, para las opciones disponibles, su variabilidad, su efectividad en base a los resultados de estudios donde se hayan comparado entre sí, su seguridad y el modo más adecuado de individualizar su uso.

El proyecto CER puede ayudar a resolver estos problemas, o al menos a encauzar su solución. Realmente se trata de una tarea compleja, como se puede deducir de la necesidad de la elaboración de normativas, formación de distintos comités y de organismos, dedicar fondos cuantiosos y cambiar muchas ideas y mentalidades obsoletas. Su futuro es difícil, y probablemente haya que cambiar algunos de los proyectos iniciados. Pero la alternativa no puede ser sentarse en un despacho y, basándose en la evaluación de las publicaciones de algunos ensayos clínicos diseñados para evaluar la eficacia de una intervención terapéutica, establecer, mediante generalizaciones, comparaciones teóricas y extrapolaciones de baja validez, cuál es la efectividad de dicha intervención, el beneficio de su uso en condiciones asistenciales reales e incluso la efectividad comparada con otras intervenciones disponibles, emitir dictámenes sobre su eficiencia y tener la arrogancia de señalar a los médicos y a los pacientes qué es lo que deben hacer en la práctica asistencial, sin disponer en demasiadas ocasiones de pruebas suficientes para apoyar las recomendaciones emitidas. Esta es una práctica que debe abandonarse lo antes posible por ser metodológica y científicamente errónea, porque induce al conformismo y a la confusión, porque en muchas ocasiones conlleva un elevado coste y riesgo, tanto para los pacientes como para el sistema sanitario, y porque sus extrapolaciones respecto al beneficio asistencial y a la relación coste-efectividad no suele tener suficiente base racionalmente sostenible ni defendible.

Después de edificar las estructuras y soportes básicos necesarios, lo más importante y fundamental es que el paradigma CER debe proponer, desarrollar y consolidar, lo antes posible, una metodología de investigación propia y adecuada que sirva para responder con fiabilidad y precisión a las preguntas planteadas. Los estudios experimentales deben seguir siendo la fuente fundamental de información CER, y para responder adecuadamente a las preguntas que requieren estudios experimentales es de vital importancia desarrollar los diseños y la infraestructura que generará información creíble y relevante del modo más rápido y económico posible. Esta investigación debe tener como base fundamental los resultados de ensayos clínicos pragmáticos comparativos, cuyos objetivos primarios estén dirigidos a evaluar resultados de primer nivel de interés asistencial y que incluyan el interés de los pacientes, y con una validez interna y externa suficientes.

Y puede y debe completarse con otros estudios observacionales bien diseñados y ejecutados, analizados con criterios científicos sólidos que aporten la validez necesaria a sus conclusiones, que abarquen un amplio abanico de la población susceptible de ser tratada con esas intervenciones, y que planteen objetivos primarios similares.

La elección del método para los estudios de efectividad comparada implica la consideración de múltiples factores. PCORI ha identificado cinco relevantes, incluyendo la validez interna, la validez externa (posibilidad de generalización), la variabilidad en los subgrupos de pacientes, la viabilidad y la urgencia temporal<sup>137</sup>. La importancia de estos factores variará en función de las preguntas y objetivos del estudio, pero las pruebas deben ser adecuadas. Bien es cierto que, por lo general, finalmente se van a precisar juicios clínicos que interpreten la ciencia basada en las pruebas. Las pruebas nunca son perfectas ni completas, y siempre hay información adicional útil, como es el caso de la MBE. Pretender utilizar sólo información perfecta, podría impedir la toma de decisiones, pero nunca debe obviarse que la probabilidad de equivocarse en una recomendación es más elevada a medida que las pruebas en que se basa son de menor calidad, y los médicos deben saber con claridad las debilidades de las recomendaciones no basadas en pruebas. Y es importante que se insista en la validez externa del estudio, preguntándose si los resultados son generalizables al conjunto de la población susceptible de ser tratada. La aplicación de la confusa jerarquía de la evidencia no debe sustituir la lectura crítica cuidadosa de los estudios.

El paradigma y el proyecto de CER pueden presentar debilidades desde el punto de vista metodológico, pragmático o económico, pero es probablemente el paso con mayor repercusión en la investigación médica desde la aprobación de las enmiendas Kefauver-Harris y los consiguientes cambios en la FDA y otras agencias reguladoras hasta el documento de armonización de Buenas Prácticas Clínicas. Refleja una decisión que probablemente no tenga ya marcha atrás, la voluntad de plasmar en la práctica asistencial el criterio largo tiempo defendido desde el punto de vista científico y metodológico de no confundir, y no aceptar las propuestas que las confunden, la eficacia con la efectividad, una distinción que desgraciadamente a veces sólo se limita al ámbito académico, sin tener en cuenta que es absolutamente prioritaria y esencial en la toma de decisiones en la práctica asistencial y en la gestión de los recursos sanitarios.

La manipulación de la información, las decisiones derivadas de análisis incorrectos por errores metodológicos o conceptuales, o por intereses espurios (sean comerciales, económicos o profesionales), pueden generar un terrible impacto negativo en la salud de la población. La CER trata de situarse en el escenario de la práctica médica habitual, y facilitar, en este escenario, la selección y priorización, con bases racionales y científicamente válidas, de las distintas opciones terapéuticas disponibles, en función de su efectividad y su eficiencia. Y conseguir una mayor accesibilidad y equidad en la población susceptible de ser tratada. Salvo un cambio dramático en la opinión pública o el clima político, CER ocupará un papel de importancia progresiva en la práctica clínica y en la política de gestión de la salud en los próximos años.

## Bibliografía

---

1. Federal Coordinating Council for Comparative Effectiveness Research. Report to the President and Congress. Washington, DC: Department of Health and Human Services; 2009. Accessed at <http://www.hhs.gov/recovery/programs/cer/cerannualrpt.pdf> on 10 January 2015
2. The American Recovery and Reinvestment Act of 2009. Accessed at <http://www.gpo.gov/fdsys/pkg/BILLS-111hrienr/pdf/BILLS-111hrienr.pdf> on 10 January 2015
3. The Comparative Effectiveness Research Act of 2009. Accessed at <http://www.gpo.gov/fdsys/pkg/BILLS-111hr2502ih/pdf/BILLS-111hr2502ih.pdf> on 10 January 2015
4. Diccionario de la lengua Española. Real Academia Española. 22ª ed, 2001. Accessed at <http://lema.rae.es/drae/?val=medicamento> on 10 February 2015
5. The rational use of drugs. Report of the Conference of Experts. Nairobi, 25-29 November 1985. World Health Organization 1987. Geneva.
6. Rice E. Dr. Frances Kelsey: Turning the Thalidomide Tragedy into Food and Drug Administration Reform. Research Paper Senior Division. Accessed at <http://www.section216.com/history/Kelsey.pdf> on 10 January 2015
7. Drug Amendments of 1962. Accessed at <http://www.gpo.gov/fdsys/pkg/STATUTE-76/pdf/STATUTE-76-Pg780.pdf> on 10 January 2015
8. Wardell WM. The US Drug Efficacy Study and its implication (DESI). Agents Actions. 1978; 8:421-2
9. Temple RJ. The effect of the Drug Regulation Reform Act of 1978 on clinical research, drug availability, and the public health. Ann N Y Acad Sci. 1981;368: 175-85
10. Directiva 2001/20/CE del Parlamento Europeo y del Consejo de 4 de abril de 2001 relativa a la aproximación de las disposiciones legales, reglamentarias y administrativas de los Estados miembros sobre la aplicación de buenas prácticas clínicas en la realización de ensayos clínicos de medicamentos de uso humano. Accessed at [http://ec.europa.eu/health/files/eudralex/vol-1/dir\\_2001\\_20/dir\\_2001\\_20\\_es.pdf](http://ec.europa.eu/health/files/eudralex/vol-1/dir_2001_20/dir_2001_20_es.pdf) on 10 January 2015
11. Ley 29/2006, de 26 de julio, de garantías y uso racional de los medicamentos y productos sanitarios. Accessed at <http://www.boe.es/buscar/pdf/2006/BOE-A-2006-13554-consolidado.pdf> on 10 January 2015
12. Reglamento (UE) no 536/2014 del Parlamento Europeo y del Consejo de 16 de abril de 2014 sobre los ensayos clínicos de medicamentos de uso humano. Accessed at <http://www.boe.es/doue/2014/158/L00001-00076.pdf> on 10 January 2015
13. Normas de Buena Práctica Clínica (CPMP/ICH/135/95). Accessed at [http://www.aemps.gob.es/industria/inspeccionBPC/docs/guia-BPC\\_octubre-2008.pdf](http://www.aemps.gob.es/industria/inspeccionBPC/docs/guia-BPC_octubre-2008.pdf) on 10 January 2015
14. Friedman LM, Furberg CD, DeMets DL. What is the question? In: Fundamentals of Clinical Trials. Ed by Friedman LM, Furberg CD and DeMets DL. John Wright, Massachusetts. 1984
15. Spilker B. Establishing Clinical Trial Objectives. In: Guide to Clinical Trials. Ed by B. Spilker. Raven Press, New York. 1991
16. Wilson MK, Karakasis K, Oza AM. Outcomes and endpoints in trials of cancer treatment: the past, present, and future Lancet Oncol. 2015;16:e32-e42.
17. Sedgwick P. Clinical trials: outcome measures. BMJ. 2015 Jan 12;350:h121
18. Whitehead L, Perkins GD, Clarey A, Haywood KL. A systematic review of the outcomes reported in cardiac arrest clinical trials: The need for a core outcome set. Resuscitation. 2014;88:150-7

19. Bowater RJ, Lilford RJ. Clinical effectiveness in cardiovascular trials in relation to the importance to the patient of the end-points measured. *J Eval Clin Pract.* 2011;17:547-53
20. Michaels AD, McKeown B, Kostal M, Vakharia KT, Jordan MV, Gerber IL, Foster E, Chatterjee K. Effects of intravenous levosimendan on human coronary vasomotor regulation, left ventricular wall stress, and myocardial oxygen uptake. *Circulation.* 2005;111:1504-9
21. Follath F, Cleland JG, Just H, Papp JG, Scholz H, Peuhkurinen K, et al; Steering Committee and Investigators of the Levosimendan Infusion versus Dobutamine (LIDO) Study. Efficacy and safety of intravenous levosimendan compared with dobutamine in severe low-output heart failure (the LIDO study): a randomised double-blind trial. *Lancet.* 2002;360:196-202
22. Mebazaa A, Nieminen MS, Packer M, Cohen-Solal A, Kleber FX, Pocock SJ, et al; SURVIVE Investigators. Levosimendan vs dobutamine for patients with acute decompensated heart failure: the SURVIVE Randomized Trial. *JAMA.* 2007;297:1883-91
23. Estacio RO, Jeffers BW, Hiatt WR, Biggerstaff SL, Gifford N, Schrier RW. The effect of nisoldipine as compared with enalapril on cardiovascular outcomes in patients with non-insulin-dependent diabetes and hypertension. *N Engl J Med.* 1998;338:645-52
24. Declaración de Helsinki de la Asociación Médica Mundial. Principios éticos para las investigaciones médicas en seres humanos. Accessed at [http://www.wma.net/es/30publications/10policies/b3/index.html.pdf?print-media-type&footer-right=\[page\]/\[toPage\]](http://www.wma.net/es/30publications/10policies/b3/index.html.pdf?print-media-type&footer-right=[page]/[toPage]) on 30 January 2015.
25. Sedgwick P, Hoope C. Placebo controlled trials. *BMJ* 2014;348:g1635
26. Wartolowska K, Judge A, Hopewell S, Collins GS, Dean BJ, Rombach I, et al. Use of placebo controls in the evaluation of surgery: systematic review. *BMJ* 2014;348:g3253
27. Armitage P. The role of randomization in clinical trials. *Stat Med.* 1982;1: 345-52
28. Altman DG. Randomisation. *BMJ* 1991;302:1481-2
29. Schulz KF, Grimes DA. Generation of allocation sequences in randomised trials: chance, not choice. *Lancet* 2002;359:515-9
30. Hróbjartsson A, Thomsen AS, Emanuelsson F, Tendal B, Hilden J, Boutron I, et al. Observer bias in randomised clinical trials with binary outcomes: systematic review of trials with both blinded and non-blinded outcome assessors. *BMJ* 2012; 344:e1119
31. Haynes B. Can it work? Does it work? Is it worth it? The testing of healthcare interventions is evolving. *BMJ* 1999;319:652-3
32. Van Staa TP, Leufkens HG, Zhang B, Smeeth L. A comparison of cost effectiveness using data from randomized trials or actual clinical practice: selective cox-2 inhibitors as an example. *PLoS Med* 2009;6:e1000194
33. Järvinen TL, Sievänen H, Kannus P, Jokihaara J, Khan KM. The true cost of pharmacological disease prevention. *BMJ.* 2011;342:d2175
34. Baños JE, Farré M. Principios de Farmacología Clínica. Bases Científicas de la Utilización de Medicamentos. Ed by J-E Baños y M Farré. Masson, Barcelona. 2002
35. Rothwell PM. External validity of randomised controlled trials: "to whom do the results of this trial apply?". *Lancet.* 2005;365:82-93
36. Dekkers OM, von Elm E, Algra A, Romijn JA, Vandenbroucke JP. How to assess the external validity of therapeutic trials: a conceptual approach. *Int J Epidemiol* 2010;39:89-94
37. Baños JE, Farré M. El ensayo clínico. In Principios de Farmacología Clínica. Bases Científicas de la Utilización de Medicamentos. Ed by J-E Baños y M Farré. Masson, Barcelona. 2002
38. Carrasco JL. El método estadístico en la investigación médica. Ed by JL Carrasco. 6ª ed. Ciencia 3. Distribución, Madrid 1995
39. Rawlins M. De testimonio: on the evidence for decisions about the use of therapeutic interventions. *Lancet.* 2008;372:2152-61

40. Rothwell PM. Commentary: External validity of results of randomized trials: disentangling a complex concept. *Int J Epidemiol*. 2010;39:94-6
41. Rothwell PM. Factors that can affect the external validity of randomised controlled trials. *PLoS Clin Trials*. 2006;1:e9.
42. Rothwell PM. Treating individuals 2. Subgroup analysis in randomised controlled trials: importance, indications, and interpretation. *Lancet*. 2005;365:176-86
43. Assmann SF, Pocock SJ, Enos LE, Kasten LE. Subgroup analyses and other (mis) uses of baseline data in clinical trials. *Lancet* 2000;355:1064-9
44. Wang R, Lagakos SW, Ware JH, Hunter DJ, Drazen JM. Statistics in medicine — reporting of subgroup analyses in clinical trials. *N Engl J Med* 2007;357:2189-94
45. Pocock SJ, Lubsen J More on subgroup analyses in clinical trials. *N Engl J Med*. 2008;358:2076
46. Cook TD, DeMets DL. Selected issues in the analysis. In *Introduction to Statistical Methods for Clinical Trials*. Ed by TD Cook, DL DeMets. Chapman & Hall/CRC, New York 2007
47. Packer M, O'Connor CM, Ghali JK, Pressler ML, Carson PE, Belkin RN, et al. Effect of amlodipine on morbidity and mortality in severe chronic heart failure. *N Engl J Med*. 1996;335:1107-14
48. Packer M, Carson P, Elkayam U, Konstam MA, Moe G, O'Connor C, et al. Effect of amlodipine on the survival of patients with severe chronic heart failure due to a nonischemic cardiomyopathy. *JACC Heart Fail*. 2013;1:308-14
49. Lagakos SW. The challenge of subgroup analyses--reporting without distorting. *N Engl J Med*. 2006;354:1667-9
50. Hasford J, Bramlage P, Koch G, Lehmacher W, Einhäupl K, Rothwell PM. Inconsistent trial assessments by the National Institute for Health and Clinical Excellence and IQWiG: standards for the performance and interpretation of subgroup analyses are needed. *J Clin Epidemiol*. 2010;63:1298-304
51. Wennberg JE Unwarranted variations in healthcare delivery: implications for academic medical centres. *BMJ*. 2002;325:961-4
52. Ottesen MM, Køber L, Jørgensen S, Torp-Pedersen C. Consequences of overutilization and underutilization of thrombolytic therapy in clinical practice. TRA-CE Study Group. TRAndolapril Cardiac Evaluation. *J Am Coll Cardiol*. 2001;37:1581-7
53. Ghosh S, Ziesmer V, Aronow WS. Underutilization of aspirin, beta blockers, angiotensin-converting enzyme inhibitors, and lipid-lowering drugs and overutilization of calcium channel blockers in older persons with coronary artery disease in an academic nursing home. *J Gerontol A Biol Sci Med Sci*. 2002;57:M398-400
54. Evidence-Based Medicine Working Group. Evidence-based medicine. A new approach to teaching the practice of medicine. *JAMA*. 1992;268:2420-5
55. Kuhn TS. La estructura de las revoluciones científicas. Fondo de cultura Económica, México, 1971. Accessed at [http://www.conductitlan.net/libros\\_y\\_lecturas\\_basicas\\_gratuitos/t\\_s\\_kuhn\\_la\\_estructura\\_de\\_las\\_revoluciones\\_cientificas.pdf](http://www.conductitlan.net/libros_y_lecturas_basicas_gratuitos/t_s_kuhn_la_estructura_de_las_revoluciones_cientificas.pdf) on 15 February 2015
56. Sackett DL, Rosenberg WM. On the need for evidence-based medicine. *J Public Health Med*. 1995;17:330-4
57. British Thoracic Society. Guidelines for management of asthma in adults: I. Chronic persistent asthma. *BMJ* 1990;301:651-3
58. Majeed A, Ferguson J, Field J. Prescribing of beta-2 agonists and inhaled steroids in England: trends between 1992 and 1998, and association with material deprivation, chronic illness and asthma mortality rates. *J Pub Health Med* 1999;21:395-400
59. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. *BMJ*. 1996;312:71-2
60. Biller-Andorno N, Lie RK, ter Meulen R. Evidence-based medicine as an instrument for rational health policy. *Health Care Anal*. 2002;10:261-75

61. McGavock H. Integrating evidence based medicine into clinical practice. *Pharmacoepidemiol Drug Saf.* 2005;14:211-2
62. Miles A, Loughlin M. Models in the balance: evidence-based medicine versus evidence-informed individualized care. *J Eval Clin Pract.* 2011;17:531-6
63. Greenhalgh T, Howick J, Maskrey N; Evidence Based Medicine Renaissance Group. Evidence based medicine: a movement in crisis? *BMJ.* 2014;348:g3725.
64. Thompson DF. Understanding financial conflicts of interest. *N Engl J Med.* 1993;329:573-6
65. Bekelman JE, Li Y, Gross CP. Scope and impact of financial conflicts of interest in biomedical research: a systematic review. *JAMA.* 2003;289:454-65
66. Bes-Rastrollo M, Schulze MB, Ruiz-Canela M, Martinez-Gonzalez MA. Financial conflicts of interest and reporting bias regarding the association between sugar-sweetened beverages and weight gain: a systematic review of systematic reviews. *PLoS Med.* 2013;10:e1001578
67. Hartling L, Chisholm A, Thomson D, Dryden DM. A descriptive analysis of overviews of reviews published between 2000 and 2011. *PLoS One.* 2012;7:e49667
68. Pieper D, Buechter R, Jerinic P, Eikermann M. Overviews of reviews often have limited rigor: a systematic review. *J Clin Epidemiol.* 2012;65:1267-73
69. Woolf SH. Practice guidelines: a new reality in medicine. I. Recent developments. *Arch Intern Med.* 1990;150:1811-8
70. Peter ED. Credibility, Cookbook Medicine, and Common Sense: Guidelines and the College. *Ann Intern Med.* 1994;120:966-8
71. Woolf SH. Practice guidelines, a new reality in medicine. II. Methods of developing guidelines. *Arch Intern Med.* 1992;152:946-52
72. Rawlins MD. 5 NICE years. *Lancet* 2005;365:904-8
73. Carthey J, Walker S, Deelchand V, Vincent C, Griffiths WH. Breaking the rules: understanding non-compliance with policies and guidelines. *BMJ.* 2011;343:d5283
74. Nolan JP, Soar J, Zideman DA, Biarent D, Bossaert LL, Deakin C, et al; ERC Guidelines Writing Group. European Resuscitation Council Guidelines for Resuscitation 2010 Section 1. Executive summary. *Resuscitation.* 2010;81:1219-76
75. Deakin CD, Nolan JP, Soar J, Sunde K, Koster RW, Smith GB, Perkins GD. European Resuscitation Council Guidelines for Resuscitation 2010. Section 4. Adult advanced life support. *Resuscitation.* 2010;81:1305-52
76. Penson PE, Ford WR, Broadley KJ. Vasopressors for cardiopulmonary resuscitation. Does pharmacological evidence support clinical practice? *Pharmacol Ther.* 2007;115:37-55
77. Jacobs IG, Finn JC, Jelinek GA, Oker HF, Thompson PL. Effect of adrenaline on survival in out-of-hospital cardiac arrest: A randomised double-blind placebo-controlled trial. *Resuscitation.* 2011;82:1138-43
78. Hagihara A, Hasegawa M, Abe T, Nagata T, Wakata Y, Miyazaki S. Prehospital epinephrine use and survival among patients with out-of-hospital cardiac arrest. *JAMA.* 2012;307:1161-8
79. McCartney M. Adrenaline in cardiac arrest: it's unethical for patients not to know. *BMJ.* 2014;349:g5258
80. Beckett NS, Peters R, Fletcher AE, Staessen JA, Liu L, Dumitrascu D, et al. HYVET Study Group. Treatment of hypertension in patients 80 years of age or older. *N Engl J Med.* 2008;358:1887-98
81. Mancia G, Fagard R, Narkiewicz K, Redón J, Zanchetti A, Böhm M, et al. Task Force Members. 2013 ESH/ESC Guidelines for the management of arterial hypertension: the Task Force for the management of arterial hypertension of the European Society of Hypertension (ESH) and of the European Society of Cardiology (ESC). *J Hypertens.* 2013;31:1281-357
82. Stone NJ, Robinson J, Lichtenstein AH, Bairey Merz CN, Blum CB, Eckel RH, et al. 2013 ACC/AHA guideline on the treatment of blood cholesterol to reduce

- atherosclerotic cardiovascular risk in adults: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *Circulation*. 2014;129 (Suppl 2):S1-45.
83. National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III). Third report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III) final report. *Circulation* 2002;106:3143-421
  84. Grundy SM, Cleeman JI, Merz CN, Brewer HB Jr, Clark LT, Hunninghake DB, et al. Implications of recent clinical trials for the National Cholesterol Education Program Adult Treatment Panel III guidelines. *Circulation* 2004;110:227-39
  85. Krumholz HM. Target cardiovascular risk rather than cholesterol concentration. *BMJ*. 2013;347:f7110
  86. Scandinavian Simvastatin Survival Study Group. Randomised trial of cholesterol lowering in 4444 patients with coronary heart disease: the Scandinavian Simvastatin Survival Study (4S). *Lancet*. 1994;344:1383-9
  87. Taylor AJ, Villines TC, Stanek EJ, Devine PJ, Griffen L, Miller M, Weissman NJ, Turco M Extended-release niacin or ezetimibe and carotid intima-media thickness. *N Engl J Med*. 2009;361:2113-22
  88. HPS2-THRIVE Collaborative Group, Landray MJ, Haynes R, Hopewell JC, Parish S, Aung T, Tomson J, et al. Effects of extended-release niacin with laropirant in high-risk patients. *N Engl J Med*. 2014;371:203-12
  89. Barter PJ, Caulfield M, Eriksson M, Grundy SM, Kastelein JJ, Komajda M, et al. ILLUMINATE Investigators Effects of torcetrapib in patients at high risk for coronary events. *N Engl J Med*. 2007;357:2109-22
  90. Cannon CP, Shah S, Dansky HM, Davidson M, Brinton EA, Gotto AM, et al. Determining the Efficacy and Tolerability Investigators. Safety of anacetrapib in patients with or at high risk for coronary heart disease. *N Engl J Med*. 2010;363:2406-15
  91. AASLD/IDSA/IAS-USA Hepatitis C Guidance. Accessed at <http://www.hepatitisc.uw.edu/go/treatment-infection/treatment-genotype-1/aasld-idsa-iasusa-hepatitis-c-guidance> on 20 september 2014
  92. Center for Evidence-based Policy Oregon Health & Science University. Sofosbuvir for the Treatment of Hepatitis C and Evaluation of the 2014 American Association for the Study of Liver Diseases Treatment Guidelines. Accessed at [http://www.ohsu.edu/xd/research/centers-institutes/evidence-based-policy-center/med/upload/Sofosbuvir\\_for\\_HepatitisC\\_FINAL\\_5\\_19\\_2014.pdf](http://www.ohsu.edu/xd/research/centers-institutes/evidence-based-policy-center/med/upload/Sofosbuvir_for_HepatitisC_FINAL_5_19_2014.pdf) on 20 september 2014
  93. FDA News Release. FDA approves first combination pill to treat hepatitis C. Accessed at <http://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm418365.htm> on 10 january 2015
  94. Feuerstein JD, Akbari M, Gifford AE, Hurley CM, Leffler DA, Sheth SG, Cheifetz AS. Systematic analysis underlying the quality of the scientific evidence and conflicts of interest in interventional medicine subspecialty guidelines. *Mayo Clin Proc*. 2014;89:16-24
  95. Harbour R, Miller J. A new system for grading recommendations in evidence based guidelines. *BMJ*. 2001;323:334-6
  96. Rawlins MD. Evidence, values, and decision making. *Int J Technol Assess Health Care*. 2014;30:233-8
  97. Schlander M. The use of cost-effectiveness by the National Institute for Health and Clinical Excellence (NICE): no(t yet an) exemplar of a deliberative process. *J Med Ethics*. 2008;34:534-9
  98. Russell LB, Gold MR, Siegel JE, Daniels N, Weinstein MC. The role of cost-effectiveness analysis in health and medicine. Panel on Cost-Effectiveness in Health and Medicine. *JAMA*. 1996;276:1172-7
  99. Marko NF, Weil RJ. An introduction to comparative effectiveness research. *Neurosurgery*. 2012;70:425-34

100. Manchikanti L, Falco FJ, Boswell MV, Hirsch JA. Facts, fallacies and politics of comparative effectiveness research: Part I. Basic considerations. *Pain Physician* 2010;13:E23-54
101. Järvinen TL, Sievänen H, Kannus P, Jokihaara J, Khan KM. The true cost of pharmacological disease prevention. *BMJ*. 2011;342:d2175
102. Borgstrom F, Strom O, Coelho J, Johansson H, Oden A, McCloskey EV, et al. The cost-effectiveness of risedronate in the UK for the management of osteoporosis using the FRAX. *Osteoporos Int* 2010;21:495-505
103. Sawka AM, Ismaila N, Cranney A, Thabane L, Kastner M, Gafni A, et al. A scoping review of strategies for the prevention of hip fracture in elderly nursing home residents. *PLoS One*. 2010;5:e9515
104. Rawlins MD. A population approach to the rational use of therapeutic interventions. *Clin Ther*. 2013;35:1634-8
105. European Medicines Agency. The European paediatric initiative: History of the Paediatric Regulation. EMEA/17967/04. Rev 1. Accessed at [http://www.ema.europa.eu/docs/en\\_GB/document\\_library/Other/2009/09/WC500003693.pdf](http://www.ema.europa.eu/docs/en_GB/document_library/Other/2009/09/WC500003693.pdf) on 15 February 2015
106. Committee on Comparative Effectiveness Research Prioritization. Institute of Medicine. Initial National Priorities for Comparative Effectiveness Research: Report Brief. 2009. Accessed at <http://www.iom.edu/~media/Files/Report%20Files/2009/ComparativeEffectivenessResearchPriorities/CER%20report%20brief%2008-13-09.ashx> on 15 February 2015
107. American College of Physicians. Information on cost-effectiveness: an essential product of a national comparative effectiveness program. *Ann Intern Med*. 2008;148:956-61
108. Conway PH, Clancy C. Comparative-effectiveness research—implications of the Federal Coordinating Council's report. *N Engl J Med*. 2009;361:328-30
109. Chalkidou K, Tunis S, Lopert R, Rochaix L, Sawicki PT, Nasser M, Xerri B. Comparative effectiveness research and evidence-based health policy: experience from four countries. *Milbank Q*. 2009;87:339-67
110. Faden RR, Chalkidou K. Determining the value of drugs-The evolving British experience. *N Engl J Med*. 2011;364:1289-91
111. Fairbrother G, O'Brien E, Pradhananga R, Chalkidou K. Improving Quality and Efficiency in Health Care through Comparative Effectiveness Analyses: An International Perspective. 2014 Accessed at <http://academyhealth.org/files/publications/2014CERImprovingQuality.pdf> on 15 February 2015
112. Kristensen FB. EUnetHTA and health policy-making in Europe. *Eurohealth* 2006;12:36-8
113. REGLAMENTO (CE) No 726/2004 DEL PARLAMENTO EUROPEO Y DEL CONSEJO de 31 de marzo de 2004 por el que se establecen procedimientos comunitarios para la autorización y el control de los medicamentos de uso humano y veterinario y por el que se crea la Agencia Europea de Medicamentos. Accessed at <http://www.boe.es/doue/2004/136/L00001-00033.pdf> on 15 February 2015
114. Sorenson C1, Naci H, Cylus J, Mossialos E. Evidence of comparative efficacy should have a formal role in European drug approvals. *BMJ*. 2011;343:d4849
115. European Medicines Agency. Draft Policy 0070: Publication and access to clinical trial data. 2013. Accessed at [http://www.ema.europa.eu/docs/en\\_GB/document\\_library/Other/2013/06/WC500144730.pdf](http://www.ema.europa.eu/docs/en_GB/document_library/Other/2013/06/WC500144730.pdf) on 15 February 2015
116. Steinbrook R. The European Medicines Agency and the brave new world of access to clinical trial data. *JAMA Intern Med*. 2013;173:373-4
117. Wieseler B, McGauran N, Kaiser T. New EMA policy-further measures needed to support comparative effectiveness assessments. *JAMA Intern Med*. 2014;174:1874-5
118. Ley 29/2006, de 26 de julio, de garantías y uso racional de los medicamentos y productos sanitarios. Texto consolidado. Última modificación: 30 de diciembre de 2014. Accessed at <http://www.boe.es/buscar/pdf/2006/BOE-A-2006-13554-consolidado.pdf> on 15 February 2015

119. Agencia Española de Medicamentos y Productos Sanitarios. Ministerio de Sanidad, Servicios Sociales e Igualdad. Propuesta de colaboración para la elaboración de los informes de posicionamiento terapéutico de los medicamentos Documento aprobado por la Comisión Permanente de Farmacia del SNS 21 de Mayo de 2013. Accessed at <http://www.aemps.gob.es/medicamentosUsoHumano/informesPublicos/docs/propuesta-colaboracion-informes-posicionamiento-terapeutico.pdf> on 15 February 2015
120. American College of Physicians. Information on cost-effectiveness: an essential product of a national comparative effectiveness program. *Ann Intern Med.* 2008; 148:956-61
121. Naik AD, Petersen LA. The neglected purpose of comparative-effectiveness research. *N Engl J Med.* 2009;360:1929-31
122. Schneeweiss S, Seeger JD, Jackson JW, Smith SR. Methods for comparative effectiveness research/patient-centered outcomes research: from efficacy to effectiveness. *J Clin Epidemiol.* 2013;66(8 Suppl):S1-4
123. Luce BR, Kramer JM, Goodman SN, et al. Rethinking randomized clinical trials for comparative effectiveness research: the need for transformational change. *Ann Intern Med.* 2009;151:206-9
124. Pocock SJ, Gersh BJ. Do current clinical trials meet society's needs?: a critical review of recent evidence. *J Am Coll Cardiol.* 2014;64:1615-28
125. Schwartz D, Lellouch J. Explanatory and pragmatic attitudes in therapeutical trials. *J Clin Epidemiol.* 2009;62:499-505
126. Mullins CD, Whicher D, Reese ES, Tunis S. Generating Evidence for Comparative Effectiveness Research Using More Pragmatic Randomized Controlled Trials. *Pharmacoeconomics* 2010; 28: 969-76
127. Sedgwick P. Explanatory trials versus pragmatic trials. *BMJ.* 2014;349:g6694
128. Chalkidou K, Tunis S, Whicher D, Fowler R, Zwarenstein M. The role for pragmatic randomized controlled trials (pRCTs) in comparative effectiveness research. *Clin Trials.* 2012;9:436-46
129. Moher D, Schulz KF, Altman D; CONSORT Group (Consolidated Standards of Reporting Trials). The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. *JAMA.* 2001;285:1987-91
130. Zwarenstein M, Treweek S, Gagnier JJ, Altman DG, Tunis S, Haynes B, et al; CONSORT group; Pragmatic Trials in Healthcare (Practihc) group. Improving the reporting of pragmatic trials: an extension of the CONSORT statement. *BMJ.* 2008;337:a2390
131. Thorpe KE, Zwarenstein M, Oxman AD, Treweek S, Furberg CD, Altman DG, et al. A pragmatic-explanatory continuum indicator summary (PRECIS): a tool to help trial designers. *CMAJ.* 2009;180:E47-57
132. Prieto L, Herranz I. La inferencia Bayesiana. En: Bioestadística sin dificultades matemáticas. Análisis estadístico de datos en investigación médica y sociológica. Ed by L Prieto y I Herranz. Díaz de Santos, Madrid. 2010
133. U.S. Food and Drug Administration. Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials. Final Guidance 2010. Accessed at <http://www.fda.gov/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm071072.htm> on 15 February 2015
134. Berry DA. Bayesian approaches for comparative effectiveness research. *Clin Trials.* 2012;9:37-47
135. Connor JT1, Luce BR, Broglio KR, Ishak KJ, Mullins CD, Vanness DJ, et al. Do Bayesian adaptive trials offer advantages for comparative effectiveness research? Protocol for the RE-ADAPT study. *Clin Trials.* 2013;10:807-27
136. Gajewski BJ, Berry SM, Quintana M, Pasnoor M, Dimachkie M, Herbelin L, Barohn R. Building efficient comparative effectiveness trials through adaptive designs, utility functions, and accrual rate optimization: finding the sweet spot. *Stat Med.* 2015;34:1134-49

137. Armstrong K. Methods in comparative effectiveness research. *J Clin Oncol*. 2012;30:4208-14
138. Orden SAS/3470/2009, de 16 de diciembre, por la que se publican las directrices sobre estudios posautorización de tipo observacional para medicamentos de uso humano. Accessed at <http://www.boe.es/boe/dias/2009/12/25/pdfs/BOE-A-2009-20817.pdf> on 15 February 2015
139. Concato J, Lawler EV, Lew RA, Gaziano JM, Aslan M, Huang GD. Observational methods in comparative effectiveness research. *Am J Med*. 2010;123(12 Suppl 1):e16-23
140. Sedgwick P. Bias in observational study designs: prospective cohort studies. *BMJ*. 2014;349:g7731
141. Sedgwick P. Bias in observational study designs: cross sectional studies. *BMJ*. 2015;350:h1286
142. Mendelsohn ME, Karas RH. The protective effects of estrogen on the cardiovascular system. *N Engl J Med*. 1999;340:1801-11
143. Anderson GL, Limacher M, Assaf AR, Bassford T, Beresford SA, Black H, et al. Women's Health Initiative Steering Committee. Effects of conjugated equine estrogen in postmenopausal women with hysterectomy: the Women's Health Initiative randomized controlled trial. *JAMA*. 2004;291:1701-12
144. Ad Hoc Subcommittee of the Liaison Committee of the World Health Organisation and the International Society of Hypertension. Effects of calcium antagonists on the risks of coronary heart disease, cancer and bleeding. *Hypertens Res*. 1997;20:61-73
145. Muñoz AM, Giraldo ND, Jaimes F. Uso de variables instrumentales en investigación médica. *IATREIA* 2013;26:371-9
146. Von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP; Iniciativa STROBE. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Rev Esp Salud Publica*. 2008;82:251-9
147. Dreyer NA, Velentgas P, Westrich K, Dubois R. The GRACE checklist for rating the quality of observational studies of comparative effectiveness: a tale of hope and caution. *J Manag Care Spec Pharm*. 2014;20:301-8
148. Gliklich R, Dreyer N, Leavy M, eds. *Registries for Evaluating Patient Outcomes: A User's Guide*. Third edition. Two volumes. (Prepared by the Outcome DEcIDE Center [Outcome Sciences, Inc., a Quintiles company] under Contract No. 290 2005 00351 TO7.) AHRQ Publication No. 13(14)-EHC111. Rockville, MD: Agency for Healthcare Research and Quality. April 2014. Accessed at <http://effectivehealthcare.ahrq.gov/ehc/products/420/1897/registries-guide-3rd-edition-vol-1-140430.pdf> on 15 February 2015
149. Fitzpatrick R, Davey C, Buxton MJ, Jones DR. Evaluating patient-based outcome measures for use in clinical trials. *Health Technol Assess*. 1998;2:1-74
150. NHS Information Centre. Patient Reported Outcome Programme. Accessed at <http://www.hscic.gov.uk/proms> on 15 February 2015
151. FDA. Guidance for Industry Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. Accessed at <http://www.fda.gov/downloads/Drugs/Guidances/UCM193282.pdf> on 15 February 2015
152. Fehnel S, DeMuro C, McLeod L, Coon C, Gnanasakthy A. US FDA patient-reported outcome guidance: great expectations and unintended consequences. *Expert Rev Pharmacoecon Outcomes Res*. 2013;13:441-6
153. Patient-Centered Outcomes Research Institute. Accessed at <http://www.pcori.org/content/why-pcori-was-created> on 15 February 2015

## Sobre el autor

---

### José Francisco Horga de la Parte

LICENCIATURA: Medicina y Cirugía Universidad Valladolid 1975. DOCTORADO: Medicina y Cirugía Universidad Valladolid 1976. MÉDICO ESPECIALISTA EN FARMACOLOGÍA CLÍNICA: Hospital “La Paz” Madrid, 1979-1981 Hospitales en los que he tenido plaza: La Paz, de Madrid; Valdecilla, de Santander; Los Lirios, de Alcoy; General, de Alicante. Universidades en cuya Facultad de Medicina he tenido plaza: Autónoma de Madrid; Complutense de Madrid; Santander; Alicante; Miguel Hernández. SITUACIÓN ACTUAL: CENTRO: Univ. Miguel Hernández (Elche) /Consellería Sanidad Comunidad Valenciana; DEPARTAMENTO: Departamento de Farmacología, Pediatría y Química Orgánica; CATEGORÍA PROFESIONAL Y FECHA DE INICIO: Catedrático de Universidad, 1985; SITUACIÓN: Plantilla; DEDICACIÓN: A tiempo completo, con plaza asistencial vinculada como Jefe de Sección de Farmacología Clínica en el Hospital General Universitario de Alicante desde 1993; N2 de sexenios de investigación reconocidos: 6; N° de quinquenios docentes reconocidos: 6 PUBLICACIONES en revistas con índice de impacto, libros o informes técnicos: > 100 LÍNEAS DE TRABAJO: Dolor y opioides; Sistema Nervioso Simpático Ensayos Clínicos; Farmacocinética; Farmacovigilancia y Farmacoepidemiología.

OTROS ASPECTOS DE LA EXPERIENCIA PROFESIONAL Supervisor de Instalaciones Radiactivas Junto Energía Nuclear. (1982); Coordinador del Programa de Farmacovigilancia de la Comunidad Valenciana (1993-hoy); Coordinador del Programa de Ensayos Clínicos de la Comunidad Valenciana (IPECME)(1994-hoy); Responsable de la realización de informes sobre medicamentos en la Comunidad Valenciana; Presidente del Comité Ético de Investigación Clínica del Hospital General Universitario de Alicante (HGUA), y miembro de otros 7 CEIC de la Comunidad Valenciana; Miembro del Equipo Asesor de la Comisión del Parlamento Europeo para la reforma de la normativa sobre Ensayos Clínicos en la Unión Europea 1996; Miembro del Comité de Expertos de Vacunación de la Comunidad Valenciana (1996-hoy); Coordinador de Formación Médica Continuada del Colegio de Médicos de Alicante; Miembro de la Comisión Asesora de la Consellería de Sanidad para la Evaluación de los Estudios Post-autorización Observacionales Prospectivos; Director de Investigación de la Fundación para la Investigación del HGUA 2007-2012; Presidente de la Comisión de Investigación del HGUA (2007-2013); Director Unidad de Ensayos Clínicos-Instituto de Bioingeniería-Universidad Miguel Hernández (2006-hoy); Jefe de la Unidad de Ensayos Clínicos-Hospital General Universitario de Alicante; Miembro del Comité de Guía Farmacoterapéutica de la Consellería de Sanidad; Miembro y presidente del Comité Evaluador de Reacciones Adversas Comunidad Valenciana; Miembro del Consejo Científico del Instituto para la Acreditación y Evaluación de las Prácticas Sanitarias de la Consellería de Sanidad Comunidad Valenciana; Miembro del Consejo Asesor de Investigación en Ciencias de la Salud de la Consellería de Sanidad Comunidad Valenciana; Revisor experto de Diversas Revistas Científicas; Evaluador del Instituto de Salud Carlos III; Director del Departamento de Farmacología, Pediatría y Química Orgánica UMH (2011-hoy)











ESINFU0043 Sept 2015