



2

## **RESULTADOS COMUNICADOS POR LOS CLÍNICOS**

ANDRÉS GARCÍA PALOMO, ANA LÓPEZ GONZÁLEZ y  
MARÍA ROSARIO FERNÁNDEZ GONZÁLEZ



## 2.1. INTRODUCCIÓN

Invertir recursos en nuevas intervenciones sanitarias solo está justificado si la supuesta innovación llega cargada de valor. Según Porter, valor no es ni un ideal abstracto, ni una clave para reducir costes, sino el marco en el que se debe definir la mejora de la atención sanitaria y el impulso innovador del sistema. Valor es una razón con un numerador, que representa los resultados de la intervención, y un denominador con los costes que induce su implementación (Porter, 2010). (Figura 1).

Figura 1.

### VALOR COMO RAZÓN ENTRE RESULTADOS (BENEFICIOS) Y COSTES (GASTO O INVERSIÓN)

**El objetivo de un Sistema de Salud Público es devolver VALOR al paciente y que éste lo perciba.**



Porter ME. N Engl J Med. 2010

Mientras que los costes se miden en unidades monetarias, los resultados son representados en términos de beneficio incremental. Se entiende como beneficio incremental cualquier efecto favorable en uno o varios aspectos significativos para el paciente, que pueden ser: vivir más, vivir con menos síntomas o retener una funcionalidad adecuada para sus necesidades diarias, comparado con lo que había o tenía. El efecto beneficioso debe ser un aspecto de la salud o la funcionalidad diaria afectado por la enfermedad, que al paciente le preocupa y sobre el que tiene expectativas de mejorar, no empeorar o que podrá ser prevenido por efecto de la intervención (Walton *et al.*, 2015).

El cálculo de los costes no es excesivamente confuso, aunque no está exento de dificultades técnicas y metodológicas. Sin embargo, la medición del beneficio, dovela del valor, se afronta con ambigüedad, es equívoco y ampliamente incomprendido. Muchas son las razones y no pocas las dificultades.

La primera y quizá la mas importante, se encuentre en la frase "...en uno o varios aspectos significativos para el paciente", porque no todos los resultados que resultan ser beneficiosos, son significativos para ellos. No cabe duda de que la involucración de los pacientes en el proceso de desarrollo de nuevos fármacos o intervenciones innovadoras, ayudará al diseño de una investigación con objetivos relevantes para ellos. Sin embargo, una interpretación literal, rigurosa y excluyente de la frase, puede representar una enorme barrera para el progreso terapéutico. La falta de "significado para los pacientes" no puede interpretarse radicalmente como una ausencia de beneficio, porque las expectativas de aquellos son heterogéneas, cambiantes, manipulables, frecuentemente inefables y algunas veces poco realistas (Coyne y Wyrwich, 2015).

La segunda se refiere a la jerarquía de los beneficios. La supervivencia, la funcionalidad y los síntomas pueden ser etiquetados en diferentes niveles según su relevancia para los pacientes. En el superior, y probablemente el más relevante e influyente sobre los inferiores, está la supervivencia (curación) y todos sus derivados "evento-tiempo". Es posible que se trate de un objetivo primordial para la gran mayoría de los pacientes, independientemente de la patología que sufran. Sin embargo, su maximización y sacralización puede no ser realista en algunas condiciones o no deseable como objetivo único para determinados grupos de pacientes. El segundo nivel es el grado de salud alcanzado, recuperado o retenido y que se mide mediante indicadores que analizan la funcionalidad para la actividad diaria básica, instrumental, avanzada o laboral. Finalmente el tercer nivel está representado por la sostenibilidad del estado de salud alcanzado y sus indicadores suelen ser referidos a recaídas y síntomas. Es posible que los niveles 2º y 3º puedan ser intercambiables en determinados marcos y condiciones (Porter y Teisberg, 2006).

La tercera está relacionada con las características de los resultados en salud. Estos suelen ser dependientes, inherentes y específicos de la condición médica que miden. Ningún resultado en particular captura todo el beneficio de la acción terapéutica. Es mas, el valor de la asistencia sanitaria, sobre todo en pacientes con enfermedades crónicas y/o pluripatología, se alcanza progresivamente en la evolución temporal y espacial del ciclo de atención y dependen del esfuerzo coordinado de todas y cada una de las intervenciones. Por tanto, la evaluación de resultados en salud debe interpretarse siempre como la medición interrelacionada de varios indicadores informados desde varias perspectivas, los cuales posteriormente serán integrados, proporcionalmente a su relevancia, en una única medida multidimensional, que representará el beneficio percibido y clínicamente significativo para el paciente (Porter, 2008).

## 2.2. INVESTIGACION EN RESULTADOS EN SALUD (IRS)

La IRS constituye la fuente mas importante de conocimiento sobre el valor de las intervenciones sanitarias, aportando información, fundamentalmente sobre los beneficios de estas, aunque también de los costes. Constituye la base conceptual de la Medicina Basada en la Evidencia (MBE) (Badia y Lizán, 2002).

Cuando se realiza en condiciones ideales, controladas y aleatorizadas, genera información de los resultados en clave de eficacia. Si se desarrolla sobre estudios observacionales posautorización no controlados y/o de fuentes de datos e información del mundo real (Datos de la Vida Real [DVR] o *Real Word Data* [RWD], en ingles), proporciona información sobre la efectividad, concepto este que complementa y da sentido a la eficacia, lo que nos acerca mucho más al valor (Badia y Del Llano, 2000). Dado que la estructura organizativa de los sistemas de salud y sus medios de información están, por ahora muy lejos de cubrir con calidad y fiabilidad la necesidad ética de entrega de información sobre efectividad, la IRS se nutre casi exclusivamente de resultados de eficacia, lo que da una idea solo aproximada de los beneficios.

Quizá la forma mas importante de IRS sobre eficacia sea la investigación experimental biomédica con fármacos, conocida como Ensayo Clínico Aleatorizado (ECA), en el que se realizan “evaluaciones de resultado”, entendidas como cualquier efecto inducido por un fármaco o fármacos en la modificación de un evento considerado y previamente definido. Las evaluaciones de resultados deben seguir una distribución de probabilidad conocida, que facilite su detección en la población estudiada. La base de cualquier ECA es rechazar la “hipótesis nula”, es decir, que el fármaco experimental A no mejora los efectos que consigue el control B. En función de que sea acertado o fallido el rechazo de la hipótesis nula se pueden cometer dos tipos de errores, el tipo I y el II (*Figura 2: Hipótesis nula y errores tipo I y II*). Para ello se usan herramientas estadísticas ad hoc que comparan los efectos sobre el evento medido en el grupo experimental frente al control (Laporte, 1993).

El resultado obtenido en un ECA y que se usa para interpretar si el efecto de la intervención ofrece beneficio incremental, se denomina *endpoint*. Se componen de una o varias evaluaciones de resultados preespecificadas en diferentes momentos del estudio, que se miden y analizan mediante herramientas estadísticas especifi-

Tabla 1.

### HIPÓTESIS NULA Y TIPO DE ERRORES EN LA INVESTIGACIÓN BIOMÉDICA

¿Se rechaza la hipótesis NULA?	Hipótesis NULA verdadera	Hipótesis NULA falsa
Si	ERROR TIPO I (falso positivo)	CORRECTO (verdadero positivo)
No	CORRECTO (verdadero negativo)	ERROR TIPO II (Falso negativo)

Fuente: Modificado de: <https://www.eupati.eu/es/desarrollo-y-ensayos-clinicos/la-estadistica-en-los-ensayos-clinicos-conceptos-clave/>

cas. Pueden clasificarse en primarios o secundarios. Los *endpoint* primarios son los que directamente contestan a la pregunta del ECA sobre beneficio incremental. Por ello, condiciona el tamaño de la muestra y define la potencia del ECA para detectar diferencias que rechacen esa hipótesis nula ya citada. Los *endpoints* secundarios complementan y matizan esa respuesta. Este principio constituye la base para la gestión de las innovaciones farmacológicas por parte de las agencias evaluadoras y es el punto de partida para la puesta en el mercado de un fármaco. (Bakhai, Chhabra y Wang, 2006).

### ■ 2.2.1. Resultados para el desarrollo de fármacos

Las variables de resultados que definen beneficio pueden ser de evento-tiempo referidas a cuánto sobreviven los pacientes, evaluaciones clínicas que contestan a preguntas sobre funcionalidad y síntomas, o biomarcadores.

Se define biomarcador como una evaluación del paciente no influenciada por el juicio del evaluador, o la motivación o voluntad del paciente. Son ejemplos de biomarcadores las mediciones bioquímicas sanguíneas, muestras biológicas o las mediciones cuantitativas de las imágenes radiológicas. Siempre ofrecen una valoración indirecta y subrogada del beneficio. Las evaluaciones clínicas por el contrario son aquellas que se ven afectadas por la volición y/o motivación del paciente o por el juicio clínico del evaluador. Por tanto, son susceptibles de variación no relacionada con el efecto beneficioso o perjudicial de la intervención. Los COA (*Clinical Outcome Assessment*) son los instrumentos usados para medir estas evaluaciones clínicas en el contexto de los ECA<sup>1</sup>

Es indudable el éxito alcanzado por la IRS basada en los ECA. Sin embargo, cada vez son más las voces que abogan por un cambio cuantitativo y cualitativo en la investigación biomédica.

La IRS basada en las ECA sufre de una profunda, crónica e inveterada falta de atención a la elección de *endpoints* primarios y secundarios causando un gran dispendio en tiempo, recursos y conocimientos. Debido a la rigidez y miopía de las normas regulatorias sobre las ECA, la mayoría de los diseños abusan de variables duras y unidimensionales como las “evento-tiempo”, que no capturan de manera integral el beneficio real y clínicamente significativo de la intervención. Es más, muchas de ellas vienen siendo sustituidas, por numerosas razones entre las que destacan los motivos económicos, por formas subrogadas de dichas variables o directamente por biomarcadores. La ya citada ausencia de pacientes en el diseño de la investigación biomédica, la falta de sensibilidad de los diseñadores, el interés económico de algunos promotores y la excesiva variabilidad, incoherencia, heterogeneidad y falta de estandarización de la mayoría de los COA, ha dado como

<sup>1</sup> <https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM230597.pdf>. Último acceso: 27.07.2018.

resultado que lo estadísticamente significativo no deje paso a lo clínicamente relevante y que la medida del beneficio sea unidimensional. Como ya se ha señalado, el beneficio es inherentemente específico de la condición que se mide y es siempre multidimensionales, por lo que ningún resultado en particular captura toda la información sobre el beneficio.

Desde los años setenta se vienen produciendo numerosas iniciativas dirigidas a cambiar esta tendencia, mediante el desarrollo, mas allá de las evento-tiempo, de indicadores de resultados clínicos estandarizados y específicos para cada área de salud o problema sanitario. Junto a los COA (Benjamin, Vernon y Patrick, 2017) ya citados, iniciativa del ISPOR (International Society for Pharmacoeconomics and Outcomes Research), destacan los *Core Outcome Sets (COS)* (Tunis *et al.*, 2016) impulsados por el PCORI (*Patients-Centered Outcomes Research*). Aunque los datos de EMA no son demasiado claros en este sentido<sup>2</sup> FDA (US FDA, 1992) advierte de la necesidad de que los COA se incorporen de forma multidimensional a los ECA y sean parte inherente del proceso de análisis de resultados.

### ■ 2.2.1.1. Variables evento-tiempo y medidas de impacto

En la IRS basada en las ECA, las variables “evento-tiempo” se consideran variables “duras” para la definición de beneficio, porque informa sin ambages y directamente de él. La supervivencia se encuentra, casi siempre, en el primer nivel de relevancia de las expectativas de los pacientes. Esta consideración las hace insustituibles. Las variables “evento-tiempo” son reportadas por el clínico o el equipo investigador, pero no se afectan por el juicio e interpretación de éstos. No son por tanto, ni COA ni biomarcador.

Tanto EMA<sup>3</sup> como FDA<sup>4</sup> considera que la variable “evento-tiempo” de mayor relevancia es la *supervivencia global (OS)*, entendida como el tiempo desde la autorización (si es en el contexto de un ECA) o desde la inscripción en el estudio (fuera del contexto de los ECA) hasta la muerte por cualquier causa. Es la medida directa y universalmente aceptada de beneficio por su sencillez de interpretación. Es cualitativa, nominal y dicotómica. Sin embargo, tiene algunas dificultades. Encarece y enlentece el estudio al requerir una muestra poblacional mas grande y un seguimiento muy largo, sobre todo cuando las presumibles diferencias pueden ser muy pequeñas; es posible que incluya muertes no relacionadas con el proceso que se mide y, en caso de terapias encadenadas o secuenciales, el cruzamiento de brazos puede complicar la medición. Con la intención de reducir los tiempos de seguimiento, la OS está siendo expresada en periodos de tiempo, como la OS a corto

<sup>2</sup> [http://www.ema.europa.eu/docs/en\\_GB/document\\_library/Scientific\\_guideline/2017/11/WC500238764.pdf](http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2017/11/WC500238764.pdf). Último acceso: 17.07.2018.

<sup>3</sup> [http://www.ema.europa.eu/docs/en\\_GB/document\\_library/Scientific\\_guideline/2013/01/WC500137126.pdf](http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2013/01/WC500137126.pdf)

<sup>4</sup> <https://www.fda.gov/downloads/drugsGuidanceComplianceRegulatoryInformation/Guidance/UCM071590.pdf>

(OS-1 año por ejemplo), medio (OS-3 años, por ejemplo) o largo plazo (OS-5 o 10 años, por ejemplo) o como un porcentaje de pacientes vivos cuando se alcanza un periodo de seguimiento dado (porcentaje de pacientes vivos a 1 año de seguimiento, por ejemplo). Los métodos estadísticos actuariales facilitan estos cálculos.

Los costes de los ECA han forzado a EMA<sup>5</sup> y FDA<sup>6</sup> a considerar la posibilidad de analizar los beneficios de un tratamiento mediante variables subrogadas. Tanto una como la otra definen variable subrogada como una medida de laboratorio, un signo físico o una variable evento-tiempo que sustituye a una variable clínicamente significativa y directa sobre lo que siente un paciente, sus funciones o su supervivencia y que se espera que prediga indirectamente el efecto de la terapia sobre la variable directa (US FDA, 1992). Existen dos variables evento-tiempo subrogadas, la *supervivencia libre de enfermedad o progresión* (SLE-SLP) y el *tiempo al fallo del tratamiento* (TTF). Se subrogan a OS porque se entiende que cuanto más se tarde en fracasar con el tratamiento, más OS se consigue a la postre. Las dos tienen en común el abaratamiento de costes y el acortamiento de las necesidades de seguimiento. La SLP o SLE, al igual que la OS, mide el tiempo hasta el fracaso del tratamiento interpretado como la aparición de un evento, previamente definido, que se interpreta como un fallo a aquel. La SLP o SLE incluye también la muerte por cualquier causa, si esta ocurre antes que dicho evento definido. Es también una variable cualitativa, nominal y dicotómica. La TTF es igual que la SLP-SLE, pero añade como evento de fallo la suspensión del tratamiento por toxicidad. Dado que la TTF mezcla eficacia y toxicidad, tanto EMA como FDA autorizan solo el uso de SLP como subrogado de OS. No obstante está sujeta a numerosos y peligrosos sesgos, no han demostrado ser universalmente subrogada de OS y requiere tiempos de evaluación equilibrados entre los brazos de tratamiento. Por ello la SLP parece convertirse poco a poco en una variable para aprobaciones condicionadas a demostrar OS.

Para evaluar las diferencias de efecto entre los brazos experimental y control se usan las denominadas medidas de impacto, herramientas estadísticas que permiten el contraste de hipótesis y el rechazo de la hipótesis nula, base de la investigación experimental. En el campo de los ECA, son medidas de impacto las mediciones de diferencias, la reducción de riesgos (relativo y absoluto) y el número necesario a tratar (NNT) (Molina Arias y Ochoa Sangrador, 2015).

Las *mediciones de diferencia* reflejan la diferencia existente entre las variables de resultados, medidas en el grupo experimental y las del control, e indican la contribución de la intervención. La cuantía de la diferencia es la magnitud del impacto. Las diferencias de medias son la herramienta para comparar variables

<sup>5</sup> [http://www.ema.europa.eu/docs/en\\_GB/document\\_library/Scientific\\_guideline/2013/01/WC500137126.pdf](http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2013/01/WC500137126.pdf)

<sup>6</sup> <https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidance/UCM071590.pdf>

cuantitativas, mientras que las diferencias de medianas se usa para el análisis de las variables cualitativas (Molina Arias y Ochoa Sangrador, 2015).

La *reducción de riesgos*, absoluta o RAR y relativa o RRR, establece la magnitud en la que se reduce el riesgo del evento medido. La absoluta o RAR es la resta directa de los riesgos en el brazo control y experimental, mientras que la RRR es la RAR dividida por el riesgo del grupo control. Aunque es moneda de cambio común usar la RRR, esta magnifica el efecto y no cuantifica claramente el esfuerzo para conseguir el beneficio. Con un ejemplo se entiende. Así si tenemos un riesgo de mortalidad del 20 % en el grupo control y del 10 % en el intervenido, la RAR será del 10 % ( $0,20 - 0,10 = 0,10$ ), mientras que la RRR será del 50 % ( $0,10/0,20 = 0,5$ ). Ahora si el control tiene un 0,2 % y el experimental es de 0,1 %, la RAR es del 0,1 % mientras que la RRR sigue siendo del 50 % ( $0,01/0,02 = 0,5$ ) (Molina Arias y Ochoa Sangrador, 2015).

El *NNT* es la medida de impacto mas importante porque establece cuántos pacientes debemos someter a los efectos del fármaco para alcanzar beneficiar a uno. Se estima mediante el RAR y es su inversa, es decir,  $1/RAR$ . Con el ejemplo anterior, el esfuerzo en el primer caso con un RAR del 10 % es de 10 ( $1/0,10$ ), mientras que en el segundo es de 1000 ( $1/0,001$ ), con RRR semejantes. No todos los ECA aportan esta información tan relevante.

### ■ 2.2.1.2. Herramientas para evaluaciones clínicas (COA)

La FDA introdujo el término COA en octubre de 2011 para reflejar la importancia de la fuente informante en la medición de resultados clínicos. Los COA excluyen a las variables evento-tiempo porque su medición e información están influenciadas por elecciones humanas, juicios, sean estos profesionales o no, motivación y cooperación del informante. También excluyen a los biomarcadores porque estos dependen de algoritmos y procesos automatizados y no de informantes humanos. En definitiva, son mediciones basadas en evaluaciones humanas mediante observación o instrumentos de medición usados por un paciente, un clínico experto o un observador (Patrick y Burke, 2013).

Se describen cuatro tipos de COA. Las *evaluaciones de resultados de rendimiento* o *PerfO* (*Performance Outcome*) instruyen a un paciente para realizar una tarea definida que se cuantifica mediante una medición. Aunque no hay juicio clínico involucrado, es un COA porque precisa de la volición del paciente para realizar la prueba. El *resultado informado por un clínico* o *ClinRO* (*Clinician-reported Outcome*) es la evaluación del estado de salud o funcionalidad de un paciente mediante la observación, interpretación e información de la medición por un médico. El *resultado informado por un observador* u *ObsRO* (*Observer-reported Outcome*) es una evaluación que se usa cuando el paciente no puede autoreportar conceptos, observables o no. Los ObsRO proporcionan observaciones, pero no interpretaciones porque

el informante no es un profesional. Finalmente, los resultados observados por los pacientes o *PRO (Patient-reported Outcome)* son más desarrollados y conocidos. El PRO es un informe reportado por el paciente sobre su estado general, aspectos particulares o eventos relacionados con su patología o su salud, sin interpretación por parte de profesionales u observadores. Los PRO más significativos incluyen signos y síntomas, calidad de vida y funcionalidad, aunque también satisfacción y, en algunos casos, adherencia.

Existe un quinto tipo de COA, no globalmente aceptado que es el ProxyRO, un concepto mezcla de ObsRO y PRO, mucho más cercano a este último, porque es otra persona la que informa como si fuera el paciente.

Los COA proporcionan evidencia complementaria de los beneficios del tratamiento, pero en algunas ocasiones pueden considerarse como *endpoint* primarios en función de su relevancia. Los PerfO tienen una larga tradición en Medicina y en IRS, a las que últimamente se han incorporado los PRO, pero están sujetos a una rigurosa evaluación de sus aportaciones. Los ObsRO se han usado en condiciones muy específicas, como Psiquiatría o Pediatría, sobre todo, pero la consideración sistemática, sus fundamentos conceptuales y sus propiedades de medición son relativamente nuevas. Finalmente los ClinRO tienen una larga historia y son asumidos como válidos porque están cargados con el “juicio clínico”; sin embargo, carecen en muchos casos de evaluaciones rigurosas de su valor y la validez de su aportación.

Todas las iniciativas están de acuerdo en que constituyen un conjunto de información que deben integrarse y estandarizarse en los ECA y que la normativa sobre los mismos debe modificarse para recogerlos como parte fundamental y obligatoria para el investigador y los promotores. Este proceso es fundamental para incorporar los resultados centrados en el paciente (*Patient centered Outcomes* o *PCO*) en la IRS<sup>7</sup>.

### ■ 2.3. RESULTADOS REPORTADOS POR CLINICOS (CLINRO)

Según la definición de la FDA, se basa en un informe emitido por un profesional de la salud capacitado, después de la observación del paciente o la aplicación de un instrumento de medida o una escala. Implica siempre la emisión de un juicio clínico que interpreta signos observables, pruebas diagnósticas, comportamientos u otras manifestaciones físicas que se cree que están relacionadas con la enfermedad o la condición medida. Las medidas ClinRO no evalúan directamente síntomas, ya que estos solo pueden ser reportados por el paciente, salvo que ese síntoma tenga reflejo en un signo clínico (por ejemplo, palpitations que se reflejan en un ECG anormal, o fiebre que se termometra) (Powers *et al.*, 2017).

<sup>7</sup> <http://wayback.archive-it.org/7993/20180424212148/https://www.fda.gov/Drugs/DevelopmentApprovalProcess/DrugDevelopmentToolsQualificationProgram/ucm370262.htm#ClinRO&xcust=pdfmyurl>. Último acceso: 27.07.2018.

### ■ 2.3.1. Tipos de ClinRO

Según los informes del ISPOR<sup>8</sup>, las ClinRO se pueden dividir en tres tipos, lecturas e interpretaciones de pruebas (*ClinRO Reading*), calificaciones impulsadas por protocolo (*ClinRo rating*) y puntuaciones globales (*ClinRo global assessment*).

En cualquiera de las ClinRO citadas, la participación del paciente puede ser pasiva como en las ObsRO, cuando se le observa a él o a una prueba médica, o activa, como las PRO o las PerFO, según se le pregunte o se le haga realizar algún tipo de prueba. Tras ello, el clínico deberá obligatoriamente analizar los datos obtenidos, para en un segundo término informar aplicando su interpretación o “juicio clínico”, mediante una variable de resultado. Esta es la clave para diferenciar las ClinRO del resto de las COA, PerFO, PRO u ObsRO en las que no cabe el juicio clínico.

#### ■ 2.3.1.1. ClinRO Readings

Son lecturas de signos clínicos o interpretación de pruebas. En ambos casos, y tras ser observados, caracterizados y analizados, serán informados en forma de variable dicotómica (sí/no; presente/ausente) tras el juicio clínico experto o el uso de un protocolo estandarizado de evaluación. La participación del paciente no es importante para el proceso de desarrollo, análisis e interpretación.

Pueden considerarse ClinRO la medición de constantes vitales, la detección de signos exploratorios o cualquier *end point* que sea mediado por el juicio clínico y conlleve una acción dicotómica. Por ejemplo, trasfundir o no, administrar oxígeno o no, realización de un procedimiento o prueba radiológica o la hospitalización. En todo caso, las *ClinRO Readings* no suelen ser consideradas *end points* primarios en los ECA.

#### ■ 2.3.1.2. ClinRO rating

Las *ClinRO rating* o calificaciones impulsadas por tablas y protocolos, están representadas por escalas compuestas de variables categóricas, ordenadas o no, o bien medidas continuas en las que son reconocibles al menos, tres niveles posibles, que generan puntajes para representar el concepto de interés o variable de resultado.

Las *ClinRO rating* son las más abundantes. Son ejemplos en Oncología las escalas de toxicidad CTCAE<sup>9</sup> o las escalas de medición de la respuesta como los

<sup>8</sup> <https://www.fda.gov/downloads/drugsGuidanceComplianceRegulatoryInformation/Guidance/UCM071590.pdf>

<sup>9</sup> <http://www.hrc.govt.nz/sites/default/files/CTCAE%20manual%20-%20DMCC.pdf>. Último acceso: 28.07.2018.

criterios RECIST<sup>10</sup>, ejemplo palmario con el que se construyen dos variables de resultados: las *tasas de respuesta* (ORR) y la *duración de la respuesta* (DoR). La ORR es una medida de proporción e incluye la cantidad de pacientes que reducen su carga de la enfermedad, medida por un criterio previamente definido y estandarizado como el RECIST, con respecto al total de pacientes incluidos en el brazo de estudio. Derivada de ella es la DoR, entendida como la duración de la ORR en el tiempo, desde que se detecta hasta que desaparece. Aunque ni la ORR ni la DoR deben considerarse una ClinRO propiamente dicha, son la consecuencia palmaria de la aplicación del juicio clínico del experto tras la aplicación de una escala protocolizada.

Algunas *ClinRO rating* pueden transformarse en escalas dicotomizadas, cuando las puntuaciones extremas se consideran rangos dicotómicos sobre la base de criterios de respuesta definidos. De hecho, las escalas RECIST se convierten en dicotómicas cuando se interpretan, porque identifica pacientes en progresión frente a pacientes sin progresión (estabilización, respuesta parcial o respuesta completa).

En determinadas ocasiones, las ClinRo Rating pueden ser consideradas *endpoint* primarios, como ocurre en algunos fases I en Oncología, como pruebas de concepto y eficacia directa del fármaco.

Figura 2.

## CONCEPTOS DE COMPLIANCE Y PERSISTENCE EN EL CUMPLIMIENTO TERAPÉUTICO



<sup>10</sup> <http://www.irrecist.com/recist/> Último acceso 28.07.2018. Último acceso: 27.07.2018.

### ■ 2.3.1.3. *ClinRO Global Assessment*

Las *ClinRo Global Assessment*, *CGA*, incluyen impresiones globales clínicas y de cambio evolutivo y están basadas en el juicio global del clínico sobre el estado de salud del paciente o un aspecto relevante del mismo para el que las variables examinadas no están bien definidas. Las *CGA* evalúan conceptos no explícitamente definidos, dan información suplementaria. Por ello, y aunque son de utilidad en la práctica clínica real, nunca deben ser considerados *endpoint* de un ensayo clínico.

Un ejemplo de este tipo de *ClinRO* son la escala ECOG-PSy la *Comprehensive Geriatric Assessment*<sup>11</sup>.

### ■ 2.3.2. Otros *ClinRO* no agrupables en las categorías anteriores

El cumplimiento terapéutico afecta al beneficio terapéutico (Simpson, Eurich y Majumdar, 2006). Se define como el grado en que la conducta del paciente coincide con las recomendaciones médicas. El término ha sido sustituido recientemente por el de adherencia, para describir el comportamiento del paciente que toma los fármacos correctamente, a las dosis adecuadas y con la frecuencia precisa.

Para ISPOR (Lakdawalla *et al.*, 2018) cumplimiento (*compliance*) y adherencia (*adherence*) son sinónimos y significan “la medida en que un paciente actúa de acuerdo con las recomendaciones dadas por su equipo médico con respecto al cumplimiento de la dosis, intervalos y horarios del fármaco prescrito”. Define también el concepto de persistencia como la duración temporal del cumplimiento o adherencia (Figura 2).

Según ISPOR, tanto el cumplimiento como la persistencia deben ser variables de evaluación en los ECA, sobre todo cuando la medicación de estudio y/o de control es oral o autoadministrada sin la supervisión del equipo investigador. El cumplimiento se analiza mediante la razón entre la dosis tomada y la prescrita. Cuando se analiza en estudios retrospectivos se usa el concepto de tasa de posesión de la medicación (*medication possession ratio*) y es el “número de dosis dispensadas durante el periodo de dispensación”, asumiendo que el cumplimiento se produce automáticamente con la dispensación. La persistencia se define desde la fecha de inicio de tratamiento y la fecha final de la observación. Puede definirse como una variable continua o dicotómica.

Los métodos de medición de la adherencia y la persistencia en el tratamiento pueden ser los COA, según se utilice un método de evaluación u otros. Los métodos directos de evaluación son la observación médica de la toma o el

<sup>11</sup> <http://siog.org/content/comprehensive-geriatric-assessment-cga-older-patient-cancer>

uso de biomarcadores. La primera puede asimilarse como ClinRO de tipo reading porque hay un clínico o farmacéutico que observa, analiza, evalúa y obtiene un juicio clínico con una variable dicotómica (si la tomó/no la tomó). Los métodos indirectos son los COA de tipo ClinRO o PRO. Son los más sencillos y económicos, pero pecan de poca objetividad. Son ClinRO la relación entre citas y cumplimiento-persistencia (si no asiste es que tampoco cumple), la valoración de la eficacia terapéutica (si hay respuesta es porque hay cumplimiento-persistencia), el uso de marcadores fisiológicos (efectos secundarios), el recuento directo de los comprimidos que devuelve o la utilización de cuestionarios. Entre los métodos indirectos de tipo PRO se cuentan el test de Haynes & Sackett, la comprobación fingida o los diarios de paciente. En todo caso, ningún método ha demostrado ser claramente eficaz. Los ECA suelen usar el método indirecto de recuento de comprimidos.

## ■ 2.4. PROCESO DE CONSTRUCCIÓN DE UN COA<sup>12</sup>

La IRS ofrece numerosas fuentes de información de los beneficios y riesgos potenciales de las intervenciones sanitarias. Pero es un elemento fundamental garantizar la calidad de cualquier medición que pueda ser base de toma de decisiones. La calidad de cualquier medición tiene dos dominios, la validez, o grado en el que realmente se mide el fenómeno de interés, y la fiabilidad o reproducibilidad en múltiples mediciones.

Los resultados medidos en los ECA pueden ser variables evento-tiempo, biomarcadores y COA. La primera, entendida como la OS es una medida directa y objetiva cargada de validez y reproducibilidad. Los segundos son también objetivos, pero son mediciones indirectas o subrogadas, cuya reproducibilidad puede estar asegurada pero no así su validez. Finalmente los COA, pueden ser medidas directas o indirectas y generalmente subjetivas porque su información puede verse modificada por juicios, volición o motivación del informador.

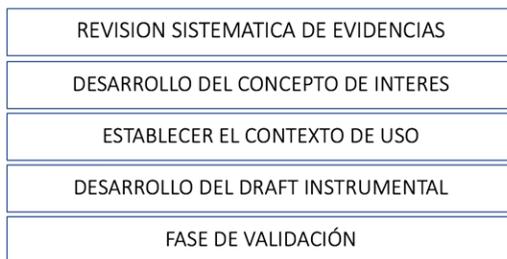
La FDA están llevando a cabo intensos y denodados esfuerzos para categorizar y estandarizar estas medidas de resultados. Fruto de ello ha sido la publicación de varios documentos tendentes a disminuir esa subjetividad, pero se han volcado en las medidas PRO<sup>13</sup>. Mientras, las ClinRO adolecen de guías que clarifiquen su desarrollo, implementación y regulación. Afortunadamente, el proceso de desarrollo de un ClinRO es prácticamente similar al que se ha diseñado para el desarrollo de los PRO, por lo que el desarrollo de constructo de los primeros favorecerá indudablemente a la eclosión de los segundos (Figura 3).

<sup>12</sup> Draft Guidance for Industry—Qualification Process for Drug Development Tools. <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM230597.pdf>

<sup>13</sup> Ibid.

Figura 3.

## FASES DE DESARROLLO DE UN COA



El proceso de construcción de cualquier COA comienza con la búsqueda de evidencias desde un enfoque sistemático para determinar su marco conceptual y aproximarse a la validez de uso, estandarte de trabajo. Es de extrema importancia que un COA que vaya a ser usado como *end point* describa directamente o pueda demostrar que tiene relación con el beneficio del tratamiento. Las fuentes de evidencia podrán proceder de revisiones bibliográficas, consultas con expertos clínicos y de entrevistas con pacientes.

El segundo paso es el desarrollo del concepto de interés (COI). La mayoría de las veces, sobre todo cuando queremos medir funcionalidad, abarcar todas las fases o tareas de una actividad concreta y medible dificulta el desarrollo de una herramienta. Dicha actividad puede deconstruirse en varias tareas e identificar una que sea corolario fácilmente medible de la actividad global. Esa tarea es el concepto de interés, sobre el que se trabajará. La experiencia del dolor es paradigmática para entender lo que es el concepto de interés. Puede diferir en numerosas facetas, intensidad, duración, frecuencia, calidad. El concepto de interés es cuando elegimos una de ellas para medirlo y desarrollamos una escala para ello.

El tercer paso es la definición del contexto de uso<sup>14</sup>: enfermedad de estudio, subpoblación de pacientes en las que puede ser aplicado, factores culturales y lingüísticos que pueden afectar a su universalización, tipo de *end point*, entorno de medición y método de administración de la medición.

El cuarto paso es ya la elaboración y desarrollo del *draft* instrumental. Es extremadamente importante definir el contexto de uso, las definiciones y descripciones de los síntomas, a menudo discrepantes entre práctica clínica y contexto

<sup>14</sup> <http://www.fda.gov/Drugs/DevelopmentApprovalProcess/DrugDevelopmentToolsQualificationProgram/ucm284077.htm>

ECA, y concretar las respuestas de cada ítem individual y sus escalas de relevancia y significación. Como el proceso es complicado e iterativo, con múltiples ciclos de revisión, es muy recomendable la formación de un grupo de expertos clínicos que reconozcan el concepto de interés medible, la definición del síntoma a medir, los dominios posibles, las escalas de valoración, el tipo de variable con el que se va a expresar la medición y el diseño del material de entrenamiento y capacitación para que la evaluación de resultados sea confiable.

Una vez diseñado, el COA pasa a la fase de validación, que depende de su fiabilidad y de su validez (Sackett, 1992). Mientras que esta expresa el grado en el que se mide el fenómeno de interés. La primera indica hasta que punto se obtienen los mismos resultados al efectuar la medición en más de una ocasión, bajo condiciones similares. Una medida precisa no tiene por qué ser válida. Así que las dos características tienen que ser analizadas. En los estudios que validan una medida se usa la comparación con un estándar ya establecido. Cuando el problema es la fiabilidad se valoran otros aspectos. La repetibilidad indica hasta que punto un instrumento proporciona resultados similares cuando se aplica a una persona de forma repetida y en idénticas condiciones. La concordancia intraobservador analiza el grado de consistencia al efectuar la medición por el mismo observador de forma repetida sobre el mismo objeto de medida y en las mismas condiciones; mientras, la consistencia interobservador mide la consistencia de la medición del instrumento cuando lo usan dos observadores distintos sobre el mismo objeto en las mismas condiciones. La fiabilidad y la validez son conceptos claves para un Clin RO exitosos.

## ■ 2.5. ATRIBUTOS DE UN COA<sup>15</sup>

Cualquier COA tiene tres atributos que deben definirse. El primero atributo es definir si el COA depende de la participación activa del paciente o del juicio del clínico. En el primero de los casos estamos ante un PRO y en el segundo ante un ClinRO. Es importante recordar que los ClinRO proporcionan la perspectiva del clínico sobre los PRO y, por tanto, a la hora de diseñar un ClinRO esta perspectiva no debe olvidarse. ¿Es necesario un ClinRO o sirve con un PRO? Si el paciente puede informar sobre su experiencia, sirve con un PRO, pero si el clínico puede “observar” el fenómeno, al PRO habrá que añadirle un ClinRO que complemente la experiencia.

El segundo atributo que debe aplicarse es definir quién es el evaluador. Hay tres categorías de evaluador, el paciente, el clínico o un observador no clínico. Debido a la diferente perspectiva, capacitación, habilidades y antecedentes, sus juicios pueden diferir y tener impacto en la calificación final.

Finalmente el tercer atributo es establecer si el COA puede medir directa o indirectamente la función o los síntomas. Los COA de medida directa proporcionan

<sup>15</sup> Ver Walton *et al.* (2015).

información sobre un aspecto significativo de forma intrínsecamente interpretable. Es directamente información Mientras los COA indirectos requieren una elaboración posterior porque aportan solo datos que deben transmutarse en información, ya que la evaluación no es inherentemente clara. Son ejemplos de estos atributos, las mediciones de agudeza visual, en la que la información vertida es inherentemente clara y mide un concepto de interés muy cercano al de la vida diaria, la agudeza visual. Sin embargo un COA indirecto, por ejemplo, la fuerza isométrica del cuádriceps, no da información inherentemente clara porque no es un concepto de interés cercano a lo que el paciente hace en su vida diaria.

## ■ 2.6. CONCLUSIONES E IDEAS FINALES

- Valor es el marco en el que se define la mejora de la atención sanitaria y el impulso innovador del sistema. Es la relación entre los resultados de la intervención y el coste o inversión que induce su implementación.
- Los resultados se miden en términos de beneficio incremental, entendido como cualquier efecto favorable en uno o varios aspectos significativos para el paciente, que pueden ser vivir más, vivir con menos síntomas o retener una funcionalidad adecuada para sus necesidades diarias.
- No todo el beneficio incremental es “significativo para el paciente”. Por ello, es importante que el paciente se involucre en la definición e información de beneficios, sin que esto suponga una interpretación literal, rigurosa y excluyente de otros informantes.
- Aunque el beneficio es jerarquizable, con la supervivencia como objetivo prioritario, esta a veces puede quedar supeditada a otras ganancias más modestas objetivamente, pero más deseables subjetivamente. La retención de funcionalidad y la mejora de síntomas pueden ser más deseables en algunos contextos.
- Cualquier resultado en salud y, por tanto, su plasmación en beneficios, es dependiente, inherente y específico de la condición médica que mide. Ningún resultado en particular captura todo el beneficio global de la acción terapéutica.
- La Investigación en Resultados en Salud es la fuente principal de la Medicina Basada en la Evidencia porque informa de beneficios en términos de eficacia, cuando su información procede del Ensayo Clínico (ECA), o en términos de efectividad cuando procede del RWD o DVR.
- En el marco del ECA se entiende por “evaluaciones de resultado” a cualquier medición de los efectos inducidos por un fármaco en la modificación de un evento considerado y previamente definido, que permita rechazar la “hipótesis nula” y evitar el error tipo I.

- Las evaluaciones de resultados que se usan para interpretar beneficio, se denominan *endpoint*. Pueden clasificarse en primarios o secundarios. Los primarios contestan directamente a la pregunta sobre beneficio; los secundarios complementan y matizan esa respuesta.
- Las evaluaciones de resultados que definen beneficio pueden ser de evento-tiempo referidas a cuánto sobreviven los pacientes; evaluaciones clínicas que contestan a preguntas sobre funcionalidad y síntomas; o biomarcadores.
- Las evaluaciones clínicas se diferencian de las variables evento-tiempo y de los biomarcadores en que se ven afectadas por la volición y/o motivación del paciente o por el juicio clínico del evaluador, sea este un clínico o un observador no cualificado. Por tanto, son susceptibles de variación no relacionada con el efecto beneficioso o perjudicial de la intervención. Se denominan COA cuando se usan en el contexto de los ECA y vienen siendo impulsados por el ISPOR y la FDA.
- Se describen cuatro tipos de COA en función de quién sea el informante: PRO cuando es el paciente, ObsRO cuando es un observador no cualificado pero cercano al paciente, ClinRO cuando el informante es un clínico experto y PerFO cuando el informante no es una persona sino una escala no interpretable por el juicio clínico. Proporcionan evidencia complementaria de los beneficios del tratamiento, y algunos, en determinadas ocasiones pueden funcionar como *endpoint* primarios.
- El *resultado informado por un clínico* o ClinRO (*Clinician-Reported Outcome*) se basa en la emisión de un juicio clínico por un profesional sanitario experto y entrenado, después de la observación del paciente o la aplicación de un instrumento de medida o una escala sobre signos observables, pruebas diagnósticas, comportamientos u otras manifestaciones físicas que se cree que están relacionadas con la enfermedad o la condición medida. Tienen una larga historia en Medicina. Han sido asumidos como válidos porque están cargados con el “juicio clínico”, pero carecen, en muchos casos, de evaluaciones rigurosas de su valor y la validez de su aportación.
- ISPOR<sup>16</sup> diferencia tres ClinRO, las *Reading* (lectura de signos o interpretación de pruebas), las *Rating* (calificaciones y juicios impulsados por escalas o protocolos) y las *Clinical Global Assessment* (CGA). En cualquiera de ellas la participación del paciente puede ser pasiva o activa.
- El cumplimiento terapéutico, en cuanto a adherencia y *compliance*, afecta al beneficio terapéutico y, por tanto, deben incorporarse como COA, bien ClinRO o PRO.

<sup>16</sup> <https://www.fda.gov/downloads/drugsGuidanceComplianceRegulatoryInformation/Guidance/UCM071590.pdf>

- La construcción de un COA sigue las mismas directrices que cualquier proceso de construcción de un indicador que asegure su validez de constructo, su fiabilidad interna y externa, así como una adecuada definición de sus atributos. El desarrollo de un ClinRO debe iniciarse por la pregunta ¿es necesario?. Solo lo será si la experiencia del paciente que comunica un PRO puede ser observada e interpretada. Los ClinRO deben ser interpretados como complementarios de los PRO. Solo los sustituyen cuando lo que siente los pacientes o su funcionalidad es una experiencia inefable.

## ■ REFERENCIAS

- BADIA, X. y LIZÁN, L. (2002). Reflexiones sobre la investigación de resultados en salud. *Aten Primaria*, 30, pp. 388-91.
- BADIA, X. y DEL LLANO, J. (2000). La investigación de resultados en salud. *Med Clin (Barc)*, 114, pp. 1-7.
- BAKHAI, A., CHHABRA, A. y WANG, D. (2006). Endpoints. En: D. WANG y A. BAKHAI (Eds), *Clinical Trials. A practical guide to design, analysis and reporting* (pp. 37-46). Chicago: Remedica.
- BENJAMIN, K., VERNON, M. K. y PATRICK, D. L. (2017). Patient-Reported Outcome and Observer-Reported Outcome Assessment in Rare Disease Clinical Trials: An ISPOR COA Emerging Good Practices Task Force Report. *Value in Health*, 20(7), pp. 838-855.
- COYNE, K. S. y WYRWICH, K. W. (2015). ISPOR Task Force For Clinical Outcomes Assessment: Clinical Outcome Assessments: Conceptual Foundation—Report of The ISPOR Clinical Outcomes Assessment – Emerging Good Practices For Outcomes Research Task Force. *Value in Health*, 18(6), pp. 739-40.
- LAKDAWALLA, D. N., DOSHI, J. A., GARRISON, L. P. Jr. *et al.* (2018). Defining Elements of Value in Health Care—A Health Economics Approach: An ISPOR Special Task Force Report [3]. *Value Health*, (2), pp. 131-139.
- LAPORTE, J. R. (1993). Principios básicos de investigación clínica. *ZénecaFarma/Ergon*. Madrid.
- MOLINA ARIAS, M. y OCHOA SANGRADOR, C. (2015). Ensayo Clínico (IV). Resultados. Variables y medidas de efecto e impacto. *Evid Pediatr.*, 11, p. 33.
- PATRICK, D. L. y BURKE, L. (2013). Focusing on the patient in drug development and research. *ISPOR Connections*, 19(6), pp. 5-8.
- PORTER, M. E. (2008). *Defining and introducing value in health care. In: Evidence-based medicine and the changing nature of health care: 2007. IOM annual meeting summary* (pp. 161-172). Washington, DC: Institute of Medicine.
- . (2010). What is value in health care? *N Engl J Med.*, 363(26), pp. 2477-2481.
- PORTER, M. E. y TEISBERG, E. O. (2006). *Redefining health care: creating value-based competition on results*. Boston: Harvard Business School Press.
- POWERS, J. H. 3rd, PATRICK, D. L., WALTON, M. K. *et al.* (2017). Clinician-Reported Outcome Assessments of Treatment Benefit: Report of the ISPOR Clinical Outcome Assessment Emerging Good Practices Task Force. *Value Health*, 20(1), pp. 2-14.
- SACKETT, D. L. (1992). A primer on the precision and accuracy of the clinical examination. *JAMA*, 267, pp. 2638-2644.
- SIMPSON, S. H., EURICH, D. T., MAJUMDAR, S. R. *et al.* (2006). A meta-analysis of the association between adherence to drug therapy and mortality. *BMJ*, 333(7557), p.15.

- TUNIS, S. R., CLARKE, M., GORST, S. L. *et al.* (2016). Improving the relevance and consistency of outcomes in comparative effectiveness research. *J Comp Eff Res*, 5(2), pp.193-205.
- US FDA. (1992). New drugs, antibiotic and biological drug product regulations; accelerated approval. Proposed rule. 57 *Federal Register*, pp. 13234-13242.
- WALTON, M. K., POWERS, J. H. 3rd, HOBART, J. *et al.* (2015). Clinical Outcome Assessments: Conceptual Foundation-Report of the ISPOR Clinical Outcomes Assessment - Emerging Good Practices for Outcomes Research Task Force. International Society for Pharmacoeconomics and Outcomes Research Task Force for Clinical Outcomes Assessment. *Value in Health*, 18(6), pp.741-52.