

12

Los resultados comunicados por los pacientes
en los ensayos clínicos

Lo importante para el paciente es lo que él percibe

MICHAEL HERDMAN
NÚRIA FERNÁNDEZ

SUMARIO

| | |
|--|-----|
| 1. Introducción | 328 |
| 2. ¿Qué es un resultado comunicado por el paciente? | 328 |
| 3. ¿Por qué utilizar un resultado comunicado por el paciente? ¿Para qué sirve? | 329 |
| 4. La calidad en el uso de los resultados comunicados por el paciente en inves- tigación clínica: algunas sombras | 331 |
| 5. Las críticas: algunos comentarios | 332 |
| 5.1. Modo de administración | 332 |
| 5.2. Interpretación de resultados | 333 |
| 5.3. Uso de instrumentos adecuados y válidos | 334 |
| 5.4. Datos no registrados | 335 |
| 5.5. La presentación de resultados | 336 |
| 6. Nuevas directrices y guías: algunas luces para guiar el desarrollo y uso de los resultados comunicados por los pacientes | 337 |
| 6.1. Directrices de la Agencia de Medicamentos y Alimentos de los EEUU .. | 337 |
| 6.2. Las reflexiones de la Agencia Europea de Medicamentos | 340 |
| 7. Cómo comunicar los resultados de los ensayos clínicos con los resultados comunicados por los pacientes: la declaración CONSORT | 340 |
| 8. Conclusiones | 342 |
| Referencias | 343 |

LO QUE VD. APRENDERÁ EN ESTE CAPÍTULO

1. Los instrumentos que recogen los resultados comunicados por el paciente han cobrado una importancia cada vez mayor en los últimos años, especialmente en la investigación clínica, en estudios epidemiológicos y en las evaluaciones económicas de las intervenciones y la tecnología médica.
2. Estos instrumentos se utilizan para definir con mayor precisión cómo percibe el paciente la repercusión de la enfermedad y del tratamiento en su salud y funcionalidad. Es un término general que abarca una serie de posibles variables de medición que incluyen los síntomas, el estado de salud, la calidad de vida relacionada con la salud y la satisfacción del participante en el ensayo clínico con el tratamiento.
3. Son especialmente útiles cuando algunos efectos de la enfermedad o su tratamiento solo son conocidos por el paciente; cuando la supervivencia no es el único resultado de interés; para complementar la información sobre mejorías en la supervivencia conseguidas con un nuevo tratamiento y conocer mejor el equilibrio entre los riesgos y beneficios del nuevo tratamiento.
4. Como con cualquier tipo de metodología para medir la eficacia y la efectividad de las intervenciones médicas o de la salud pública, hay problemas e inconvenientes asociados a su uso.
5. Hacen falta directrices sobre cómo mejorar el uso de estos instrumentos así como la comunicación de resultados obtenidos con ellos. Existen documentos y guías que pueden ayudar a solventar al menos algunos de los problemas como los publicados por las agencias de regulación americana y europea y la declaración CONSORT sobre cómo redactar el informe en el caso de ensayos clínicos con asignación aleatoria.
6. Es de esperar que la publicación de estos documentos ayude a mejorar el desarrollo, aplicación y comunicación de los resultados obtenidos con estos instrumentos tan relevantes para los pacientes.

1. INTRODUCCIÓN

Los resultados comunicados por el paciente (*patient reported outcomes, PRO*) han cobrado una importancia cada vez mayor en los últimos años, especialmente en la investigación clínica,¹ en estudios epidemiológicos y en las evaluaciones económicas de las intervenciones y la tecnología médica. Aunque su papel en la atención clínica aún no está bien establecida,² los datos obtenidos con los PRO se utilizan con creciente frecuencia para informar y guiar la atención centrada en el paciente, en la toma de decisiones clínicas, en la política sanitaria o decisiones sobre el reembolso de medicamentos y productos sanitarios.

Sin embargo, quedan preguntas y dudas sobre los PRO, sobre su proceso de desarrollo, su aplicación y sobre cómo se tienen que interpretar y comunicar los resultados obtenidos con ellos. En este capítulo se pretende repasar algunos puntos claves respecto al uso de este tipo de instrumento, describir algunas de las ventajas que aporta y detallar algunos de los problemas asociados a su uso en la investigación clínica y las iniciativas para solventar esos problemas.

2. ¿QUÉ ES UN RESULTADO COMUNICADO POR EL PACIENTE?

El término PRO fue acuñado para evitar la confusión en el proceso regulador en relación con los diferentes tipos de datos obtenidos en el marco de los ensayos clínicos.³ Aunque los PRO se han utilizado en medicina durante mucho tiempo, las escalas formales empezaron a desarrollarse durante la década

de 1960, y cada vez se utilizan con más frecuencia para definir con mayor precisión cómo percibe el paciente la repercusión de la enfermedad y del tratamiento en su salud y funcionalidad.⁴

Aunque pueden existir varias definiciones de lo que es un PRO, una que ha cobrado bastante importancia en los últimos años es la que se incluyó en uno de los documentos reguladores más relevantes de los últimos años, la guía de la Agencia de Medicamentos y Alimentos de los EEUU (FDA) publicada en 2009 y titulada *Medidas comunicadas por los pacientes: su uso en el desarrollo de productos médicos para respaldar las especificaciones de la ficha técnica*.⁵ Según este documento, un PRO es cualquier expresión del estado de salud de un paciente realizada directamente por él mismo y sin que su respuesta sea interpretada por ni por un médico ni por nadie más. Esta definición concuerda bastante con la de la Agencia Europea de Medicamentos (EMA), según la cual un PRO es un resultado directamente valorado por el paciente y basado en su percepción de una enfermedad y/o su tratamiento.⁶

Por otro lado, «PRO» es un término general que abarca una serie de posibles variables de medición que incluyen los síntomas, el estado de salud, la calidad de vida relacionada con la salud (CVRS) y la satisfacción del participante en el ensayo clínico con el tratamiento. La CVRS es un tipo de evaluación de varios dominios, en general definida como una percepción subjetiva del paciente de los efectos de la enfermedad y del tratamiento en la vida diaria, el bienestar y la función fisiológica, física y social. En el

contexto de la autorización de fármacos, la CVRS se considera un tipo concreto de PRO. Más recientemente, los ensayos clínicos también han incorporado instrumentos PRO centrados en la satisfacción, el cumplimiento y las preferencias de tratamiento del paciente.

Es importante destacar que, hoy en día, se considera que los instrumentos PRO están formados no únicamente por el medio de captación de datos (v.gr., una escala, un diario o un cuestionario) sino por toda la información y documentación necesarias para su utilización, es decir, un manual de uso si existe, datos sobre su validez, fiabilidad y la sensibilidad a los cambios del instrumento, así como la información necesaria para poder interpretar los datos (valores de referencia como las normas poblacionales y/o datos sobre la diferencia clínica mínimamente importante).

De todos modos, aunque parece que ahora hay bastante acuerdo sobre lo que es un PRO, sería una negligencia no señalar el importante grado de debate generado en torno a algunos de los conceptos, la calidad de vida en particular, medidos por los PRO. Aunque entrar en detalles de ese debate va más allá del alcance de este capítulo, es importante señalar los comentarios de Sonja Hunt y Alain Leplege que criticaban algunas de las bases conceptuales de la medición de la calidad de vida en el campo de la salud, entre otras cosas por dar demasiada importancia al funcionalismo o por confundir la salud con la calidad de vida.⁷ En esta misma línea, otros autores han señalado el poco esfuerzo hecho en muchos casos para investigar la noción de la calidad de vida, entre otros

conceptos PRO, en culturas no occidentales,^{8,9} aunque se ha de reconocer el gran esfuerzo realizado por algunos grupos, especialmente los investigadores que han colaborado en el desarrollo del WHOQOL,¹⁰ para investigar más a fondo el grado de aplicabilidad universal de tales conceptos.

3. ¿POR QUÉ UTILIZAR UN RESULTADO COMUNICADO POR EL PACIENTE? ¿PARA QUÉ SIRVE?

La incorporación de los PRO a la investigación clínica ha significado el ir más allá de la utilización de indicadores puramente clínicos o epidemiológicos para medir el impacto de una enfermedad o una intervención. A través de la estandarización de la medición de conceptos relativamente abstractos como la calidad de vida, la satisfacción con el tratamiento o el bienestar emocional, entre otros, se ha conseguido ampliar nuestro concepto de la salud, tal y como proponía la Organización Mundial de la Salud, con su definición de la salud como un estado de completo bienestar físico, mental y social, y no solamente la ausencia de enfermedad o dolencia.¹¹

Los PRO son, además, muy útiles en varias situaciones que aparecen con frecuencia, sobre todo en las enfermedades crónicas (tabla 1). En primer lugar cuando algunos efectos de la enfermedad o su tratamiento solo son conocidos por el paciente como, por ejemplo, el dolor, el picor, la sensación de bienestar o el estigma que puede estar asociado con tener una determinada enfermedad o afección. En estos casos es improbable que un médico o un familiar, por

*Tabla 1. Situaciones en las que los resultados comunicados por los pacientes (PRO) son especialmente útiles**

1. Cuando algunos efectos de la enfermedad o su tratamiento solo son conocidos por el paciente.
2. Cuando la supervivencia no es el único resultado de interés.
3. Cuando la información aportada por los PRO complementa la información sobre mejorías en la supervivencia conseguidas con un nuevo tratamiento.
4. Cuando el uso de un PRO puede ayudar a conocer mejor el equilibrio entre los riesgos y beneficios de un nuevo tratamiento o intervención.

* sobre todo en enfermedades crónicas

ejemplo, pueda dar una evaluación muy precisa de lo que siente el paciente. En segundo lugar, cuando la supervivencia no es el único resultado de interés. Como se ha comentado, esta situación es especialmente relevante en las enfermedades crónicas, en las cuales la enfermedad y las intervenciones para manejarla no tienen un gran impacto en la expectativa de vida sino en la calidad de vida de los pacientes.¹² Aunque la medición de conceptos como la CVRS, el bienestar emocional o la función social pueden ser relevantes sobre todo en las enfermedades crónicas, también pueden serlo en las agudas, si queremos reflejar con fidelidad el impacto global de esas enfermedades. Por último, cuando la información aportada por los PRO puede complementar la información sobre mejorías en la supervivencia conseguidas con un nuevo tratamiento. Puede ser, por ejemplo, que el nuevo tratamiento aunque aporte mejorías en la supervivencia lo haga a costa de un impacto negativo en la calidad de vida. Conocer mejor este equilibrio entre los riesgos y beneficios del nuevo tratamiento puede ser muy importante para los pacientes

que tengan que tomar una decisión respecto a si aceptarán el tratamiento o no.

Por otro lado, hay bastantes pruebas de que no existe una correlación alta entre las medidas puramente fisiológicas y las medidas PRO. Las correlaciones suelen ser débiles o moderadas, lo cual sugiere que miden diferentes aspectos de un concepto central (salud) y que, por tanto, puede ser importante medir los dos aspectos para captar bien el impacto de la enfermedad y su tratamiento. En un estudio reciente, por ejemplo, se encontró una correlación de 0,46 entre el cambio en el FEV1 y el cambio en la puntuación total del Cuestionario Respiratorio Saint George en pacientes con enfermedad pulmonar obstructiva crónica.¹³ Este tipo de resultado indica que un cambio en un indicador clínico tan importante como el FEV1 no se traduce necesariamente en un cambio de magnitud similar en el estado de salud subjetivo del paciente (que, al final, es probablemente lo que más le importa).

Por último, hay que destacar que una evaluación bien desarrollada por los pacientes es tan fiable o más que las calificaciones globales de los médicos.

4. LA CALIDAD EN EL USO DE LOS RESULTADOS COMUNICADOS POR EL PACIENTE EN INVESTIGACIÓN CLÍNICA: ALGUNAS SOMBRAS

Como con cualquier tipo de metodología o tecnología para medir la eficacia y la efectividad de las intervenciones médicas o de la salud pública, hay problemas e inconvenientes asociados al uso de los PRO. Estos problemas pueden exacerbarse en el caso de los PRO por su relativa novedad. Aquí repasamos los resultados de algunos estudios que han examinado el uso de PRO en diferentes áreas de la investigación clínica, para reflejar algunas de las *sombras* más relevantes en el uso reciente y actual de los PRO.

Dinan *et al.*¹⁴ llevaron a cabo un estudio para investigar los patrones actuales de uso de los PRO en los ensayos clínicos a doble enmascaramiento, controlados con placebo y con asignación aleatoria, que utilizaron uno o más PRO entre los años 2004 y 2006. Encontraron 180 ensayos clínicos con asignación aleatoria que cumplían con los criterios de inclusión. Los autores destacaron que ninguno de los instrumentos utilizados en los 180 ensayos se administró por vía electrónica, un hallazgo que coincide con los datos proporcionados en otra revisión¹⁵ que también demostró una falta de PRO administrados por vía electrónica en los ensayos clínicos con asignación aleatoria, a pesar de su potencial para reducir limitaciones prácticas en la administración de este tipo de instrumento. Otro aspecto señalado por Dinan *et al.*¹⁴ fue que pocos de los estudios revisados describieron o mencionaron la diferencia clínica percibida por

el paciente como mínimamente importante (MID): únicamente un 11% de los ensayos identificados lo citaron.

Abbott *et al.*¹⁶ realizaron una revisión de la calidad de los ensayos clínicos con asignación aleatoria realizados entre 1991 y 2004 en pacientes con fibrosis quística y en los que se medía la calidad de vida. Identificaron 16 ensayos que cumplían con los criterios de inclusión. Entre los problemas revelados destacaron la falta de justificación para el uso de un determinado instrumento, el uso de escalas no validadas previamente para su uso en pacientes con fibrosis quística y la baja calidad de la información comunicada en algunos de los artículos. En especial, criticaron el hecho de que algunos artículos no proporcionaron ni una descripción completa de la escala ni una explicación del método de puntuación, y que, en algunos ensayos en los cuales se habían incluido niños, no siempre estaba claro quién había contestado el cuestionario (el hijo, el padre o en combinación); la tendencia a interpretar los resultados de calidad de vida en términos de significación estadística más que en términos de su significado clínico, y una gestión relativamente pobre de los datos no registrados. Como comentaron los autores, «el hecho de que una escala se haya utilizado previamente en una población con fibrosis quística, no quiere decir que se haya validado adecuadamente para su uso en estudios de fibrosis quística».¹⁶

Le Cleach *et al.*¹⁷ estudiaron la calidad de la información proporcionada en ensayos clínicos con asignación aleatoria en dermatología que habían utilizado PRO. Entre los 67 estudios recuperados,

el 87% no describió adecuadamente el instrumento o instrumentos utilizados, el 75% no abordó la gestión de los datos no registrados, y el 95% no proporcionó un cálculo del tamaño de la muestra necesario para obtener la potencia estadística deseada. Finalmente, el 95% no dio a conocer las puntuaciones medias pre- y post- intervención para los PRO utilizados.

Frampton y Shepherd¹⁸ evaluaron críticamente el uso de PRO específicos para asma en 87 ensayos clínicos con asignación aleatoria de los medicamentos por vía inhalatoria publicados durante el periodo 1985-2006. Encontraron que solo ocho (10%) de los estudios utilizaron instrumentos validados y en solo cinco (6%) se intentó proporcionar una interpretación clínica de las puntuaciones obtenidas en los instrumentos empleados. También destacaron un nivel pobre a la hora de comunicar los resultados de los estudios, así como un análisis e interpretación de datos poco crítico o adecuado. Sugirieron que hacía falta identificar y recomendar un grupo de PRO validados para uso en investigación clínica en asma, así como un documento *guía* para orientar a los investigadores sobre la aplicación, el análisis y la interpretación de PRO medidos en los ensayos clínicos.

Finalmente, Kyte *et al.*¹⁹ sugirieron que, aunque hay varias directrices disponibles para informar sobre el diseño y presentación de informes de los ensayos clínicos que incluyen PRO, hay una falta de directrices sobre cómo manejar estos instrumentos mientras el estudio está en marcha (es decir, sobre aspectos como el reclutamiento, la recopilación y la introducción de datos), y también sobre

cómo gestionar los casos de datos PRO potencialmente preocupantes (es decir, los datos que pueden llevar a la preocupación por el bienestar de los participantes del ensayo). Para comprobar si esto realmente fue así, realizaron una revisión sistemática de artículos con directrices sobre cómo manejar la implementación de PRO durante un ensayo. Localizaron 18 artículos que proporcionaban directrices de este tipo. Sin embargo, encontraron que la mayoría de las directrices (72%) revisadas se referían a actividades *preensayo*, por ejemplo, selección de medidas de evaluación y problemas de diseño, mientras un 16% trataban más con las actividades *postensayo*, como análisis de datos, presentación de informes e interpretación de resultados. Solo el 9% trataban realmente con aspectos relevantes mientras el ensayo está en marcha, aspectos como cumplimiento, control de calidad, evaluaciones por terceros cercanos al participante (*proxies*) e informes sobre la recogida de datos. No hubo directrices sobre los PRO *preocupantes* como, por ejemplo, instrumentos que contengan preguntas sobre la actividad sexual.¹⁹

5. LAS CRÍTICAS: ALGUNOS COMENTARIOS

5.1. *Modo de administración*

El tema remarcado por Dinan *et al.*¹⁴ sobre el uso de diferentes plataformas para la administración de los cuestionarios PRO es de gran relevancia dada la posibilidad de administrarlos tanto en papel, como por vía electrónica o telefónica. Llama la atención que una forma de administración que puede ayudar

a facilitar la recogida y estandarización de los datos, como es la vía electrónica, no se aplicase en ninguno de los estudios identificados, aunque esto puede haber cambiado desde los años que abarcaba la revisión (2004-2006). De todos modos, es importante tener en cuenta que el uso de diferentes modos de administración (papel y lápiz, administración electrónica o por teléfono) puede afectar a los resultados obtenidos y, si es posible, hay que investigar hasta qué punto esto puede ser cierto con diferentes instrumentos y en diferentes poblaciones. Aunque un meta-análisis de los estudios que habían examinado la equivalencia entre puntuaciones obtenidas con los cuestionarios PRO administrados en papel y electrónicamente demostró que son, en general, muy comparables,²⁰ no es necesariamente el caso con los administrados vía entrevistador o telefónicamente en comparación con la autoadministración. Por ejemplo, hay algunos estudios que indican diferencias en la puntuación según el modo de administración.^{21,22} Por tanto, es un aspecto a tener en cuenta a la hora de diseñar un estudio y a la hora de comparar los datos con los de otros estudios. Es importante también remarcar que existen recomendaciones sobre las pruebas que se tendrían que aportar para demostrar la equivalencia de medición de los cuestionarios PRO entre versiones electrónicas y en papel.²³

5.2. Interpretación de resultados

El tema de la MID mencionado en más de una de las revisiones comentadas, tiene que ver con cómo interpretamos las puntuaciones de un instrumento de

recogida de PRO expresadas en escalas ordinales o continuas, a las que no estamos acostumbrados. Incluso los que están habituados al concepto de evaluación de los PRO o la CVRS, en general no tienen nociones intuitivas del significado de un cambio en la puntuación de una magnitud determinada en la mayoría de instrumentos que los investigadores emplean. Por lo tanto, es un problema de cómo se interpretan: ¿qué cambios corresponden a un beneficio o perjuicio trivial, pequeño, moderado o grande para el paciente? Este problema se puede abordar de diferentes maneras. Una de ellas, la MID, sea beneficiosa o perjudicial, conduciría al paciente o al médico a plantearse un cambio de tratamiento.²⁴ Pero, ¿cómo podemos saber qué es la MID para un determinado instrumento? Para ello se han propuesto diferentes abordajes. Uno mediante el uso de *anclajes* o referencias externas y el otro mediante una investigación de la distribución estadística.

En el método *anclaje* se comparan los resultados en la medida PRO de interés con un anclaje o referencia externa que es en sí mismo interpretable, es decir, que tiene una relevancia conocida para los pacientes. Esto puede ser, por ejemplo, una calificación global del cambio, el estado en una importante y fácil de comprender medida de función, la presencia de síntomas, la gravedad de la enfermedad, la respuesta al tratamiento, el pronóstico de los acontecimientos futuros, o la utilización de servicios de salud.²⁴ Con el enfoque basado en la distribución de las puntuaciones, se interpretan los resultados en términos de la relación entre la magnitud del efecto

y alguna medida de la variabilidad en los resultados. La estadística más utilizada con este enfoque es el tamaño del efecto (*effect size*), junto con la interpretación de Cohen, que clasifica un efecto pequeño como 0,2, un efecto moderado como 0,5 y un efecto grande en un instrumento como 0,8.²⁴ Un ejemplo de la aplicación de los dos enfoques lo describen Pickard *et al.*²⁵ En este caso los autores querían conocer el MID para el EQ-5D, un instrumento utilizado con frecuencia, especialmente en estudios de coste-efectividad. Los autores utilizaron las puntuaciones en dos cuestionarios muy utilizados en pacientes con cáncer —el cuestionario Evaluación Funcional de Terapia del Cáncer (FACT) y el sistema de clasificación del Eastern Cancer Oncology Group (ECOG)— como puntos de anclaje para categorizar a los pacientes. Después analizaron qué cambio de puntuación en el EQ-5D correspondía a un cambio entre una categoría y otra en las dos escalas específicas. Adicionalmente, examinaron qué puntuación en el EQ-5D fue equivalente a una media desviación típica y el error estándar de medición en la población de interés. Para el índice del EQ-5D las estimaciones del MID basado en ECOG variaron desde 0,10 hasta 0,12 (en una escala de 0 a 1), tanto para todos los cánceres como para el subgrupo de cáncer de pulmón. Usando quintiles en el FACT-G, los MID fueron de 0,09 a 0,10 para todos los cánceres, y de 0,07 a 0,08 para el cáncer de pulmón (cuando se utilizaron valores del Reino Unido para calcular el índice). Las MID para la Escala Visual Analógica del EQ-5D fueron similares para todos los tipos de cáncer de pulmón e iban del 8 al

12 (ECOG) y de 7 a 10 (FACT-G). MID basados en la media del error estándar y una desviación típica de 0,5 oscilaron entre 0,08 y 0,14 (puntuaciones de Reino Unido), y de 0,07 a 0,12 (puntuaciones de los Estados Unidos).

Aunque conocer la MID de un instrumento puede ser muy útil a la hora de interpretar las puntuaciones y los cambios en las puntuaciones de un cuestionario, es también importante remarcar que la MID puede variar según la población (v.gr., en pacientes menos graves comparado con pacientes más graves, o pacientes en diferentes países) y según el enfoque utilizado. Por tanto, es recomendable utilizar varios enfoques para calcular los resultados para la población de interés.

5.3. *Uso de instrumentos adecuados y válidos*

La falta de datos sobre su validez puede ser un problema muy grave para un PRO, porque esta ausencia impide saber si el instrumento realmente mide de forma fiable lo que pretende medir. Por tanto, es fundamental comprobar las propiedades psicométricas del instrumento. Estas propiedades consisten en la validez (que los ítems y dimensiones de un instrumento realmente miden el concepto teórico subyacente), la coherencia interna (hasta qué punto todas las preguntas de un cuestionario PRO miden el concepto teórico subyacente), la reproducibilidad de los resultados de un cuestionario PRO al aplicarlo repetidamente al mismo grupo de pacientes, y la capacidad de detectar cambios clínicamente importantes a lo largo del tiempo.²⁶ La demostración

de que un instrumento es sensible a los cambios puede ser de especial relevancia en un ensayo clínico con asignación aleatoria ya que, de lo contrario, un resultado negativo podría ser un artefacto de la falta de sensibilidad de la escala y no del verdadero efecto del tratamiento sobre la calidad de vida. En general, con respecto a la validez y la sensibilidad, como comentan Abbott *et al.*¹⁶ «puede ser más ético no medir la calidad de vida en absoluto que utilizar una escala no adecuada, que podría proporcionar resultados engañosos».

Finalmente, respecto a este punto, merece la pena destacar que la «falta de idoneidad» de los cuestionarios utilizados en los estudios se citó como una de las principales razones por las que la información contenida en las propuestas de fichas técnicas basadas en PRO fue rechazada por la FDA.²⁷

5.4. Datos no registrados

En cuanto a la falta de registro de datos, hay que comentar que aunque no es posible entrar en detalle sobre este aspecto del uso de los PRO en este capítulo dado que es un tema de mucha envergadura, sí es posible destacar que se puede tratar en dos momentos bien diferenciados. Por un lado, antes de comenzar la recogida de datos y durante la realización del estudio, mediante técnicas y estrategias destinadas a reducir al máximo la cantidad de datos no registrados y, por otro lado, después de acabar el trabajo de campo, mediante la exploración de los datos y el uso de técnicas estadísticas que pretenden reducir cualquier sesgo al que pueden conducir.

Para reducir el número de datos no registrados en la recogida de PRO, Kyte *et al.*¹⁹ desarrollaron una serie de recomendaciones que incluyeron: que el investigador del estudio motive al paciente para que cumplimente todos los cuestionarios, asegurándose de que comprenda el objetivo y la importancia de la evaluación realizada mediante PRO en el estudio; que a la recepción de los cuestionarios, compruebe que estén completos y anime al paciente a contestar todas las preguntas que no haya cumplimentado; que muestre aprecio por el esfuerzo del paciente al terminar el cuestionario; que los PRO se recojan en el contexto clínico, sin distracciones, y que se aplique un sistema de seguimiento de estos datos gestionado a nivel central, para evaluar el cumplimiento en todos los puntos clave del estudio.

A nivel estadístico hay que remarcar que se han investigado y aplicado métodos de creciente sofisticación en el análisis del efecto de los datos no registrados. Así, Peyre *et al.*²⁸ compararon la puntuación media personal, información completa de máxima verosimilitud, imputación múltiple y técnicas *hot deck* (un método para el manejo de datos en el que cada valor no registrado se reemplaza con una respuesta observada desde una unidad similar) para determinar la mejor manera de tratar los datos no registrados en el SF-36 con datos de la encuesta nacional de salud francesa. Para todos los tipos de datos no registrados, la imputación múltiple y la información completa de máxima verosimilitud parecían ser superiores a la puntuación media personal y a la técnica *hot deck* en términos de su precisión. Sin embargo,

encontraron que el uso de la puntuación media personal se asoció con un nivel de sesgo insignificante en todas las situaciones estudiadas. Otro ejemplo de una exploración del efecto de los datos no registrados sobre los resultados de un estudio es el trabajo de Moinpour *et al.*²⁹ que explora el efecto de los datos no registrados y no ignorables en un ensayo clínico con asignación aleatoria de fase III de gemcitabina más paclitaxel frente a paclitaxel en monoterapia en pacientes con cáncer de mama metastático. En este caso utilizaron análisis de sensibilidad con seis modelos de mezcla de patrones longitudinales para evaluar los posibles sesgos debido al abandono, junto con gráficos de probabilidad acumulativa con intervalos de confianza del 50% para identificar la proporción de pacientes que mejoraron, se deterioraron o que permanecieron sin cambios en los dos cuestionarios utilizados en el ensayo. Aunque puede haber otros enfoques igual de válidos o más para determinar el efecto de los datos no registrados sobre los resultados finales del estudio, lo importante de este último quizás es el esfuerzo realizado para tenerlos en cuenta y la sofisticación de los métodos utilizados para tal propósito.

5.5. La presentación de resultados

Para que los PRO tengan influencia en la toma de decisiones deben considerarse valiosos y ser comprendidos tanto por médicos como por pacientes, ya que contribuyen factores de ambas partes al posible fracaso de dicha influencia. Por ejemplo, hay indicios de que los ensayos clínicos que incluyen un PRO a menudo

no pueden explicar la importancia clínica de los hallazgos, y por tanto, los médicos pueden no tener claro cómo interpretar su importancia para la práctica clínica. Asimismo, ello puede conllevar que los médicos no entiendan o no valoren los PRO y no los comuniquen a los pacientes. Incluso superando estos obstáculos, y si se comunican los PRO sistemáticamente, la toma de decisiones compartida podría verse perjudicada si los pacientes no comprenden lo suficiente la información. A este respecto, se ha observado que presentar los datos de los PRO complejos y multidimensionales en forma de gráfico, podría ser un método eficaz para aumentar la comprensión de los resultados de ensayos clínicos por parte del paciente.³⁰

Un problema importante, que surge en ocasiones, es lograr que se comprenda qué significa el resultado evaluado. Dicha comprensión supone un desafío incluso en el caso de resultados dicotómicos aparentemente sencillos, como un ictus, un infarto de miocardio o la muerte. La complejidad aumenta al darnos cuenta de que ningún resultado binario es del todo claro: la muerte puede ser dolorosa o indolora, el ictus puede ser leve o grave, y el infarto de miocardio puede ser grande y complicado, o pequeño y no complicado. La forma en que los investigadores presentan los resultados de los ensayos clínicos también influye en la voluntad de los médicos de realizar una acción concreta. Este problema se complica aún más cuando se considera que el valor asignado a un beneficio determinado puede depender del paciente (variación interindividual), o que incluso un mismo paciente puede asignar valores distintos a un mismo beneficio

(variación intraindividual) en función de las circunstancias. Estas dificultades surgen a pesar de la facilidad con la que en general puede comunicarse un acontecimiento como un ictus o la muerte.

6. NUEVAS DIRECTRICES Y GUÍAS: ALGUNAS LUCES PARA GUIAR EL DESARROLLO Y USO DE LOS RESULTADOS COMUNICADOS POR LOS PACIENTES

En esta sección queremos destacar algunas de las directrices y guías para el desarrollo y uso de los PRO más relevantes de los últimos años, ya que su propósito es intentar contrarrestar algunos de los problemas descritos hasta este momento. De nuevo, es imposible describir aquí en detalle todas las directrices y guías existentes, pero sí comentaremos brevemente algunas de las más importantes.

6.1. *Directrices de la Agencia de Medicamentos y Alimentos de los EEUU*

Estas directrices, que no son vinculantes sino que deben considerarse solo recomendaciones, describen cómo la FDA revisa y evalúa los instrumentos de recogida de PRO existentes, modificados o de reciente creación, que se utilizan para respaldar especificaciones de la ficha técnica de medicamentos aprobados. Son probablemente las directrices que más impacto han tenido en los últimos años, debido a su importancia para la industria farmacéutica a la hora de presentar una característica que se quiere incluir en la ficha técnica (*labelling claim*), cuando se ha utilizado un PRO en los ensayos clínicos con asignación aleatoria destinados a apoyar esa rei-

vindicación (*claim*). Pretenden también mejorar la comunicación entre la industria farmacéutica y la FDA respecto al uso de estos instrumentos, y hacer más eficiente la revisión por parte de la FDA de la idoneidad y los datos de los PRO recogidos durante un ensayo clínico. En la tabla 2 se resumen los aspectos más relevantes de estas directrices.

Como se puede apreciar, las directrices cubren un amplio rango de los aspectos más relevantes en el desarrollo y aplicación de los PRO. Quisiéramos destacar, sin embargo, tres aspectos claves. El primero es el énfasis que pone la FDA en la validez de contenido, es decir, el grado en que un instrumento mide los aspectos importantes que los que desarrollan el instrumento pretenden evaluar, y la implicación de los pacientes (o miembros de la población diana) en el proceso de generación de su contenido. Las directrices consideran este un aspecto de importancia fundamental. Además, hacen hincapié en la necesidad de poder demostrar de dónde procede cada ítem en un cuestionario a través de herramientas como las matrices de seguimiento.³¹ Otro aspecto destacable son las dudas de la FDA⁵ sobre el uso de un marco temporal o periodo recordatorio demasiado largo. Prefieren el uso de un periodo recordatorio corto. Como la FDA señala «los instrumentos PRO que obligan a los pacientes a depender de la memoria, especialmente si tienen que recordar un largo periodo de tiempo, comparar su estado actual con un periodo anterior, o el promedio de sus respuestas en un periodo de tiempo, tienden a socavar la validez de contenido. Además, es probable que la respuesta sea influenciada por

Tabla 2. Características de un instrumento para evaluar los resultados comunicados por los pacientes (PRO) que revisa la Agencia de Medicamentos y Alimentos de los EEUU (FDA)⁵

| | |
|--|---|
| Revisión del tipo de instrumento de recogida de PRO | Idoneidad del proceso de generación de las preguntas |
| | Número de preguntas |
| | Marco conceptual, relación de cada pregunta con las demás, y de los dominios entre ellos |
| | Trastorno médico para el que se usa |
| | Población en la que va a utilizarse |
| Evaluación de la validez del contenido | Método de recogida de datos: papel, electrónico, comparabilidad entre métodos |
| | Modo de administración: autoadministración, entrevista o una combinación |
| | Opciones de respuesta: es necesario ver si son coherentes con su objetivo y uso previsto |
| | El marco temporal del instrumento, es decir, el periodo recordatorio utilizado (el día de hoy, los últimos 7 días, etc). |
| | Puntuación: en cada pregunta en general se asignan puntuaciones numéricas a cada categoría de respuesta |
| | Importancia y fuente de las preguntas o dominios |
| | Formato: el promotor deberá tener en cuenta todas las instrucciones de los instrumentos de recogida de PRO de las publicaciones y manuales pertinentes |
| | Carga para el participante: la presión física, emocional o cognitiva en el paciente, en general, disminuye la calidad y cumplimentación completa de los datos de PRO |
| | Disponibilidad de traducción y de adaptación cultural |
| Evaluación de la fiabilidad, la validez conceptual y la capacidad del instrumento de recogida de PRO para detectar cambios que puedan afectar a las respuestas | Fiabilidad: la capacidad del instrumento de recogida de PRO de proporcionar estimaciones sistemáticas y reproducibles de los verdaderos efectos del tratamiento |
| | La validez conceptual viene determinada por la evidencia de relaciones entre preguntas, dominios y conceptos |
| | La capacidad de detectar cambios se revisará empleando datos que comparen cambios en las puntuaciones de los PRO con cambios en otras mediciones similares, que indiquen que el estado del paciente ha cambiado respecto al concepto de interés |

| | | |
|--|---|--|
| Instrumentos de recogida de PRO para segmentos de población especiales | Niños y adolescentes | Se valorará el vocabulario, la comprensión y la especificación de la edad mínima |
| | Personas discapacitadas | Se aconsejan informes de observadores |
| | Varias culturas/idiomas | Se valorarán pruebas de validez de la medición en todas las versiones |
| Aspectos que el promotor deberá evitar | La transmisión directa de datos de PRO del sistema de recogida de datos al promotor, investigador clínico u otras partes, sin pista de auditoría electrónica | |
| | Que solo el promotor controle el documento original y el investigador no pueda mantener y confirmar la precisión de los datos electrónicos de PRO | |
| | La existencia de una sola base de datos sin copia de seguridad y que nadie que no sea el investigador modifique los datos originales | |
| | La pérdida de datos sobre acontecimientos adversos y el acceso prematuro o no planificado a datos no desvelados | |
| | Que personal de la FDA no pueda inspeccionar, verificar y copiar los datos en el lugar donde se realizó el ensayo clínico durante una inspección | |
| | Un sistema inseguro de registro y la transmisión directa de datos de PRO importantes sobre seguridad al promotor, investigador, etc., sin asegurar la transmisión planificada de los datos al investigador clínico responsable de los pacientes | |
| | Escoger selectivamente a posteriori resultados de PRO usados como criterio de valoración para incluirlos en las especificaciones propuestas (en la ficha técnica) | |
| Análisis estadístico | La FDA determinará la idoneidad de los datos del ensayo clínico para respaldar especificaciones en la ficha técnica, en función del método propuesto para el análisis de los criterios de valoración | |
| | Los análisis estadísticos no planificados o realizados a posteriori, tras desvelar la asignación de los tratamientos y de forma exploratoria, no sirven como base para una especificación de eficacia en la ficha técnica. | |
| | Se anima al promotor a evitar proponer especificaciones en la ficha técnica basadas exclusivamente en la significación estadística | |

el estado del paciente en el momento de recordar».⁵ Finalmente, es interesante la insistencia de la guía en la necesidad de desarrollar un modelo conceptual de variables (*endpoints*), para ilustrar la relevancia de las diferentes posibles variables en la evaluación de un fármaco concreto, así como los comentarios sobre la dificultad de demostrar mejoría en un concepto o variable tan multidimensional como la CVRS, y su preferencia, en general, por variables más *concretas* como un síntoma o una función concreta.

6.2. *Las reflexiones de la Agencia Europea de Medicamentos*

La EMA no ha emitido todavía un documento de características similares al de la FDA antes comentado. Se ha limitado, hasta la fecha, a publicar unas reflexiones (*reflection paper*) sobre el uso de instrumentos de CVRS en la evaluación de productos medicinales. Este documento⁶ es bastante más escueto que las directrices de la FDA⁵ y no proporciona recomendaciones sobre, por ejemplo, el desarrollo de los instrumentos PRO. El enfoque es, sobre todo, comentar el lugar que un cuestionario de CVRS puede tener en el proceso de evaluación de los medicamentos, y proporciona algunas recomendaciones generales sobre su uso en el contexto de los documentos de orientación existentes. La EMA pone bastante énfasis en la necesidad de utilizar instrumentos ya validados y que sean adecuados para la afección en estudio. Destaca que para reivindicar una mejoría en la CVRS es necesario demostrar una mejoría en la mayoría de dimensiones que componen la CVRS para la

afección de interés, y que puede ser más relevante estudiar solo una o dos de las dimensiones más relevantes.⁶ Como en el caso de la FDA, la EMA insiste en la jerarquía de las variables: «El método recomendado para el control de la multiplicidad de criterios de valoración es una estructura jerárquica de criterios de valoración. La variable más importante (eficacia) se prueba primero, si es significativa, entonces se prueba la segunda variable y así sucesivamente. Si la variable principal no resulta ser significativa, no se analizan las siguientes».⁶ Finalmente, es interesante señalar que la EMA también comenta la importancia de la MID, pero lo hace en el contexto del cálculo del tamaño de la muestra más que en el contexto de la interpretación de datos. Concretamente, comenta que «la MID se puede utilizar para el cálculo del tamaño de la muestra del estudio. Se debe tener en cuenta, sin embargo, que la determinación de la MID debe basarse en una combinación de estadística, razonamiento y juicio clínicos y ninguno de ellos es suficiente por sí solo».⁶ Por tanto, también la EMA insiste en la necesidad de utilizar múltiples enfoques para estimarlo. En las tablas 3 y 4 se resumen las diferencias y similitudes entre la directriz de la FDA y las reflexiones de la EMA.

7. CÓMO COMUNICAR LOS RESULTADOS DE LOS ENSAYOS CLÍNICOS CON RESULTADOS COMUNICADOS POR LOS PACIENTES: LA DECLARACIÓN CONSORT

Como comentamos anteriormente, uno de los problemas identificados en relación al uso de los PRO era una baja calidad a la hora de informar sobre los

Tabla 3. Resultados comunicados por el paciente (PRO). Diferencias entre la directriz de la Agencia de Medicamentos y Alimentos de los EEUU (FDA)⁵ y las reflexiones de la Agencia Europea de Medicamentos (EMA)⁶

| | FDA ⁵ | EMA ⁶ |
|---|---|--|
| Recomendaciones relativas a los instrumentos de calidad de vida relacionada con la salud (CVRS) | Recomienda identificar conceptos relevantes en función de lo que se desee incluir en la ficha técnica, según el modelo de criterios de valoración y según las ventajas terapéuticas previstas de la molécula a analizar | Proporciona una <i>lista</i> de los instrumentos validados a utilizar en distintos ámbitos terapéuticos |
| | Considera que un PRO es cualquier percepción del paciente no filtrada por el médico | Distingue la CVRS de otros instrumentos de recogida de PRO multidimensionales |
| | Se considera la CVRS demasiado inespecífica y amplia como para respaldar una especificación de la ficha técnica | Diferencia la CVRS de los síntomas principales de una enfermedad que se consideran criterios de eficacia en ensayos para la autorización de comercialización |
| Periodo a recordar por parte del paciente | Aconseja emplear un diario para evaluar los síntomas | No impone este requisito |
| Empleo de PRO electrónicos | En general tiene dudas en cuanto al periodo a recordar, relacionados con el sesgo derivado de la memoria si el periodo a recordar es superior a 7 días | |
| Presentación de documentos | Tiene un modelo de <i>dossier</i> para la presentación de un PRO | No exige un formato concreto |
| Nivel de evaluación | Pretende revisar el <i>dossier</i> de PRO y evalúa si es relevante (validez del contenido, generación de preguntas, periodo a recordar, modificación de un cuestionario) y si la diferencia entre grupos es significativa | Para cualquier criterio de valoración (incluidos los PRO), pregunta si el síntoma es muy conocido, si hay publicaciones al respecto, si se ha utilizado previamente en la presentación de un <i>dossier</i> y si ya se ha utilizado en una especificación (de ficha técnica), y si la diferencia entre grupos es significativa |

Tabla 4. Resultados comunicados por el paciente (PRO). Similitudes entre la directriz de la Agencia de Medicamentos y Alimentos de los EEUU (FDA)⁵ y las reflexiones de la Agencia Europea de Medicamentos (EMA)⁶

- Modelo de variables de evaluación (*endpoint model*): jerarquía de las variables de evaluación, en la que se especifica cuáles son las principales y cuáles las secundarias.
- La importancia que se da a la validez del contenido: contenido y dominios relevantes para la población que sufre la enfermedad y sus tratamientos.
- El instrumento se valida en la población/trastorno a estudiar antes de utilizarse en ensayos para respaldar una especificación de la ficha técnica.
- Se termina la validación antes de utilizar los resultados comunicados por el paciente en ensayos de fase III.

resultados obtenidos con ellos. En un intento de mejorar la forma de proporcionar la información sobre los resultados obtenidos con estos instrumentos en ensayos clínicos, se ha publicado recientemente³² una extensión de la declaración CONSORT. Esta es una lista-guía de 22 puntos destinada a mejorar la comunicación de los resultados de los PRO utilizados en ensayos clínicos. Estas nuevas recomendaciones (tabla 5) instan a que se identifique el PRO como variable (*endpoint*) principal o secundaria en el resumen del artículo, que se proporcione una descripción de la hipótesis respecto a los instrumentos usados y las dimensiones que se consideren pertinentes, que se proporcionen pruebas sobre la validez y fiabilidad del instrumento PRO utilizado, que se expliciten los métodos estadísticos utilizados para tratar los datos no registrados, y que se comenten las limitaciones específicas de los PRO, así como la posibilidad de generalizar los resultados a otras poblaciones o a la práctica clínica.

Finalmente, creemos que vale la pena mencionar una serie de artículos publicados en los últimos años por autores vinculados a la International Society for Pharmacoeconomic and Outcomes Research sobre varios aspectos relacionados con el desarrollo y uso de los PRO. Algunos de ellos ya han sido citados, pero también vale la pena mencionar otros artículos sobre el uso de estos instrumentos en población pediátrica sobre validez de contenido^{34,35} y sobre la traducción y validación lingüística de estos instrumentos.³⁶

8. CONCLUSIONES

En este capítulo hemos intentado dar una idea de la utilidad de los PRO y de algunos de los problemas identificados con su uso en la investigación clínica, así como describir varias de las iniciativas que pueden ayudar a solventar esos problemas. No hay duda de que los PRO se utilizarán cada vez más en la investigación clínica, en los estudios epidemiológicos y, posiblemente en el

Tabla 5. Declaración CONSORT para redactar el informe de ensayos clínicos con asignación aleatoria con los resultados comunicados por el paciente (PRO)³²

| Apartado | Apartado de la lista de verificación de la declaración CONSORT 2010 | Extensiones específicas sobre los instrumentos de recogida de PRO |
|--------------------------|---|---|
| Antecedentes y objetivos | Título y resumen | En el resumen del informe debe indicarse si el PRO es una variable principal o secundaria |
| | Introducción | Se debe indicar la hipótesis del PRO y deben identificarse los dominios pertinentes, cuando se trabaje con instrumentos multidimensionales |
| Resultados | Métodos | Si están disponibles, deben proporcionarse o citarse pruebas de la validez y fiabilidad del instrumento de recogida de PRO, incluida la persona que responde del instrumento de PRO y los métodos de recogida de los datos (papel, teléfono, electrónico u otros) |
| Métodos estadísticos | Métodos estadísticos empleados para comparar grupos en cuanto a las variables principales y secundarias | Deberán indicarse explícitamente cuáles son los enfoques estadísticos aplicados en caso de que no se hayan registrado algunos datos |
| Limitaciones | Discusión | Limitaciones específicas de los PRO e implicaciones en la generalización de los resultados y a la práctica clínica |
| Interpretación | | Los datos de los PRO deben interpretarse en relación con los resultados clínicos, incluidos los datos de supervivencia, según corresponda |

futuro, en la práctica clínica. Como es un campo relativamente nuevo dentro de la investigación y de la medicina, existen problemas aún con el desarrollo y uso de estos instrumentos. Sin embargo, estamos convencidos que, dado el interés que generan y el dinamismo de este campo, los problemas acabarán so-

lucionándose y los PRO aportarán grandes beneficios a la investigación y, sobre todo, otorgarán a los pacientes una voz y una serie de oportunidades para contribuir a una investigación más acorde con lo que los pacientes necesitan y reclaman.

REFERENCIAS

1. Scoggins JF, Patrick DL. The use of patient-reported outcomes instruments in registered clinical trials: evidence from ClinicalTrials.gov. *Contemp Clin Trials*. 2009; 30: 289-92.
2. Greenhalgh J. The applications of PROs in clinical practice: what are they, do they work, and why? *Qual Life Res*. 2009; 18: 115-23.
3. Burke L. Drug Information Association workshop on pharmacoeconomics and quality of life labeling and marketing claims; 2000 Oct 3; New Orleans (LA) Horsham (PA): Drug Information Association; 2000. Acceptable evidence for pharmaceutical advertising and labeling.
4. Willke RJ, Burke LB, Erickson P. Measuring treatment impact: a review of patient-reported outcomes and other efficacy endpoints in approved product labels. *Control Clin Trials*. 2004; 25: 535-52.
5. U.S Department of Health and Human Services Food and Drug Administration Guidance for Industry. Patient-reported outcome measures: use in medical product development to support labeling claims. U.S. FDA, *Clinical/Medical*. 2009. Disponible en: <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM193282.pdf>. Consultado el 4 de julio de 2013.
6. EMA (Committee for medicinal products for human use (CMHP)). Reflection paper on the regulatory guidance for the use of health-related quality of life (HRQL) measures in the evaluation of medicinal products, 2005, Doc. Ref. EMA/CHMP/EWP/139391/2004. Londres.
7. Leplège A, Hunt S. The problem of quality of life in medicine. *JAMA*. 1997; 278: 47-50.
8. Herdman M, Fox-Rushby J, Badia X. A model of equivalence in the cultural adaptation of HRQoL instruments: the universalist approach. *Qual Life Res*. 1998; 7:323-35.
9. Bowden A, Fox-Rushby JA. A systematic and critical review of the process of translation and adaptation of generic health-related quality of life measures in Africa, Asia, Eastern Europe, the Middle East, South America. *Soc Sci Med*. 2003; 57: 1289-306.
10. Skevington SM, Bradshaw J, Saxena S. Selecting national items for the WHOQOL: conceptual and psychometric considerations. *Soc Sci Med* 1999;48:473-87.
11. Constitución de la Organización Mundial de la Salud, aprobada en la Conferencia Internacional de Salud de 1.946, y que entró en vigor el 7 de abril de 1.948. Disponible en: <http://www.who.int/about/definition/en/print.html>. Consultado el 15 de junio de 2013.
12. Badia X, Herdman M. The importance of health-related quality-of-life data in determining the value of drug therapy. *Clin Ther*. 2001; 23: 168-75.
13. Westwood M, Bourbeau J, Jones PW, Cerulli A, Capkun-Niggli G, Worthy G. Relationship between FEV1 change and patient-reported outcomes in randomised trials of inhaled bronchodilators for sta-

- ble COPD: a systematic review. *Respir Res.* 2011; 12: 40.
14. Dinan MA, Compton KL, Dhillon JK, et al. Use of patient-reported outcomes in randomized, double-blind, placebo-controlled clinical trials. *Med Care.* 2011; 49: 415-9.
 15. Jones JB, Snyder CF, Wu AW. Issues in the design of Internet-based systems for collecting patient-reported outcomes. *Qual Life Res.* 2007; 16: 1407-17.
 16. Abbott J, Hart A. Measuring and reporting quality of life outcomes in clinical trials in cystic fibrosis: a critical review. *Health Qual Life Outcomes.* 2005; 3:19.
 17. Le Cleach L, Chassany O, Levy A, Wolkenstein P, Chosidow O. Poor reporting of quality of life outcomes in dermatology randomized controlled clinical trials. *Dermatology.* 2008; 216: 46-55.
 18. Frampton GK, Shepherd J. Patient-reported outcomes in clinical trials of inhaled asthma medications: systematic review and research needs. *Qual Life Res.* 2011; 20: 343-57.
 19. Kyte DG, Draper H, Ives J, Liles C, Gheorghe A, Calvert M. Patient reported outcomes (PROs) in clinical trials: is «in-trial» guidance lacking? A systematic review. *PLoS One.* 2013; 8: e60684.
 20. Gwaltney CJ, Shields AL, Shiffman S. Equivalence of electronic and paper-and-pencil administration of patient-reported outcome measures: a meta-analytic review. *Value Health.* 2008; 11: 322-33.
 21. Garcia-Losa M, Unda M, Badia X, et al. Effect of mode of administration on I-PSS scores in a large BPH patient population. *Eur Urol.* 2001; 40: 451-7.
 22. Lyons RA, Wareham K, Lucas M, Price D, Williams J, Hutchings HA. SF-36 scores vary by method of administration: implications for study design. *J Public Health Med.* 1999; 21: 41-5.
 23. Coons SJ, Gwaltney CJ, Hays RD, et al. Recommendations on evidence needed to support measurement equivalence between electronic and paper-based patient-reported outcome (PRO) measures: ISPOR ePRO Good Research Practices Task Force report. *Value Health.* 2009; 12: 419-29
 24. Brozek JL, Guyatt GH, Schünemann HJ. How a well-grounded minimal important difference can enhance transparency of labelling claims and improve interpretation of a patient reported outcome measure. *Health Qual Life Outcomes.* 2006; 4: 69.
 25. Pickard AS, Neary MP, Cella D. Estimation of minimally important differences in EQ-5D utility and VAS scores in cancer. *Health Qual Life Outcomes.* 2007; 5: 70.
 26. Hays RD, Anderson R, Revicki D. Psychometric considerations in evaluating health-related quality of life measures. *Qual Life Res.* 1993; 2: 441-9.
 27. DeMuro C, Clark M, Mordin M, Fehnel S, Copley-Merriman C, Gnana-sakthy A. Reasons for rejection of patient-reported outcome label claims: a compilation based on a review of patient-reported outcome use among new molecular entities and biologic license applications, 2006-2010. *Value Health.* 2012; 15: 443-8.

28. Peyre H, Leplège A, Coste J. Missing data methods for dealing with missing items in quality of life questionnaires. A comparison by simulation of personal mean score, full information maximum likelihood, multiple imputation, and hot deck techniques applied to the SF-36 in the French 2003 decennial health survey. *Qual Life Res.* 2011; 20: 287-300.
29. Moinpour CM, Donaldson GW, Liepa AM, Melemed AS, O'Shaughnessy J, Albain KS. Evaluating health-related quality-of-life therapeutic effectiveness in a clinical trial with extensive nonignorable missing data and heterogeneous response: results from a phase III randomized trial of gemcitabine plus paclitaxel versus paclitaxel monotherapy in patients with metastatic breast cancer. *Qual Life Res.* 2012; 21: 765-75.
30. McNair AG, Brookes ST, Davis CR, Argyropoulos M, Blazeby JM. Communicating the results of randomized clinical trials: do patients understand multidimensional patient-reported outcomes? *J Clin Oncol.* 2010; 28: 738-43.
31. Calvert M, Blazeby J, Altman DG, Revicki DA, Moher D, Brundage MD. CONSORT PRO Group. Reporting of patient-reported outcomes in randomized trials: the CONSORT PRO extension. *JAMA.* 2013; 309: 814-22.
32. Matza LS, Patrick D, Riley AW et al. Pediatric patient-reported outcome instruments for research to support medical product labeling: Report of the ISPOR PRO good research practices for the assessment of children and adolescents task force. *Value Health.* 2013; 16; 461-79.
33. Rothman M, Burke L, Erickson P, Leidy NK, Patrick DL, Petrie CD. Use of existing patient-reported outcome (PRO) instruments and their modification: the ISPOR Good Research Practices for Evaluating and Documenting Content Validity for the Use of Existing Instruments and Their Modification PRO Task Force Report. *Value Health.* 2009; 12: 1075-83.
34. Patrick DL, Burke LB, Gwaltney CJ, et al. Content validity - Establishing and reporting the evidence in newly-developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO good research practices task force report: Part 1 - Eliciting concepts for a new PRO instrument. *Value Health.* 2011; 14; 967-77.
35. Wild D, Eremenco S, Mear I, et al. Multinational trials – Recommendations on the translations required, approaches to using the same language in different countries, and the approaches to support pooling the data: the ISPOR Patient Reported Outcomes Translation & Linguistic Validation Good Research Practices Task Force Report. *Value Health.* 2009; 12: 430-40.