

Expresiones y criterios mejorables en la redacción de informes de evaluación y posicionamiento terapéutico

ALEGRE DEL REY EJ¹, ORTEGA ESLAVA A², FÉNIX-CABALLERO S¹, FRAGA FUENTES D³, MARTÍNEZ LÓPEZ I^{4,5}, EN NOMBRE DEL GRUPO GÉNESIS-SEFH*

1 Servicio de Farmacia. Hospital Universitario Puerto Real. Puerto Real. Cádiz (España)

2 Clínica Universidad de Navarra. Pamplona. Navarra (España)

3 Subdirección General de Farmacia. Dirección General de Cartera Común de Servicios del Sistema Nacional de Salud y Farmacia. Ministerio de Sanidad. Madrid (España)

4 Coordinación Unidad de Genética y Genómica de les Illes Balears-GENIB. Palma de Mallorca (España)

5 Coordinación Unidad de Genética Molecular y Genética Clínica. Hospital Universitario Son Espases. Palma de Mallorca (España)

* Grupo Coordinador de GÉNESIS-SEFH (por orden alfabético de primer apellido): Alegre del Rey EJ, Arocas Casañ V, Clopés Estela A, Delgado Téllez de Cepeda L, Fénix Caballero S, Flores Moreno S, Fraga Fuentes MD (coordinadora), García Sánchez S, López Briz E (coordinador adjunto), Martínez López I, Martínez López de Castro N, Ortega Eslava A, Puigventós Latorre F, Ubago Pérez R, Velasco Rocas L

Fecha de recepción: 08/02/2021 - Fecha de aceptación: 08/02/2021

RESUMEN

En la revisión de borradores de informes para evaluación y posicionamiento de nuevos fármacos en un entorno multidisciplinar, se observan determinados errores de expresión o criterio que se repiten con frecuencia. Principalmente, están relacionados con la consideración de "diferencias" o tendencias no significativas, abuso de la reducción relativa del riesgo, errores en la valoración de resultados por subgrupos sin calcular la interacción estadística, confusiones en la interpretación de las comparaciones indirectas, sobreva-

loración de la relevancia clínica con variables subclínicas y afirmaciones sesgadas en el apartado de seguridad, entre otros. También se observa a menudo ambigüedad o inhibición en el posicionamiento, especialmente en situaciones de precariedad en la evidencia disponible. El presente trabajo expone de forma sintética tales errores, aclara algunos términos comunes y propone expresiones o criterios alternativos que se consideran preferibles, con el fin de ofrecer una evaluación para la toma de decisiones en beneficio de los pacientes.

Palabras clave: **Evaluación de medicamentos, escritura médica, metaanálisis en red, medición de riesgo, modelos de riesgos proporcionales.**

Expressions and criteria that could be improved in the writing of reports for drug evaluation and therapeutic positioning

SUMMARY

In the review of reports for evaluation and positioning of new drugs in a multidisciplinary setting, some usual errors of expression or criteria are observed. Most of them are related to the consideration of "differences" or non-signi-

ficant trends, abuse of the relative risk reduction, errors in the assessment of results by subgroups without calculating the statistical interaction, misinterpretation of indirect comparisons, excess in the assessment of clinical relevance with subclinical variables and

biased statements in the safety section, among others. Ambiguity or inhibition in positioning is also often observed, especially in situations of precariousness in the available evidence. This work summarizes such errors, clarifies some common terms and proposes alternative expressions or criteria that are considered preferable, in order to offer evaluations for decision-making focused on the benefit of patients.

Key words: **Drug evaluation, medical writing, network meta-analysis, risk assessment, proportional hazards model.**

En el proceso de revisión de borradores de informes de evaluación y posicionamiento de medicamentos, ya sea a nivel estatal o para comisiones de farmacia a nivel autonómico o local, se observan con frecuencia determinadas expresiones que es preciso corregir o mejorar. Reuniendo la experiencia de los miembros del grupo GENESIS-SEFH, se han recopi-

lado las más comunes y se han agrupado en expresiones relacionadas con la interpretación y presentación del análisis estadístico (Tabla 1) y cuestiones de criterio evaluador en la redacción (Tabla 2), acompañadas de propuestas de redacción alternativa. Las expresiones se han sintetizado en las siguientes categorías:

Aspectos estadísticos

1. Identificación de supuestas diferencias o “tendencias” sin significación estadística: si no hay significación estadística, la diferencia aparente es más atribuible al azar de lo aceptable, y se debe evitar expresar que el efecto es “mayor”, “menor”, así como que exista “diferencia” o “tendencia”¹.

2. Uso de la reducción relativa del riesgo (RRR) para expresar el beneficio: induce sobreestimación. Es preferible usar la reducción absoluta del riesgo (RAR) o la diferencia de respuestas, y el NNT (número necesario de pacientes tratados para que uno se beneficie), éste especialmente en intervenciones preventivas².

3. Uso de RRR en el análisis de supervivencia (ej: HR 0,75 → RRR 25%): sobrevalora y confunde. El HR obtenido en un modelo de Cox es una razón instantánea de riesgos. Además de la sobreestimación ya comentada, cuando se refiere un análisis de supervivencia tiende a confundirse la relación instantánea (HR) con la acumulada, y puede hacer creer que una “reducción de mortalidad del 25%” supone recuperar un

25% del tiempo de vida perdido por la enfermedad, o evitar un 25% de las muertes. Es preferible expresar simplemente el HR y, cuando sea adecuado, medias/medianas y diferencia.

4. Valoración de subgrupos sin interacción, haciendo análisis intragrupal y no intergrupales³: la interpretación de los resultados por subgrupos requiere revisar la p de interacción o calcularla ($p < 0,05$, significativa, o al menos $0,05 < p < 0,1$, dudosa), para comprobar si los subgrupos se comportan de forma diferente entre sí. Sin interacción, los resultados de los subgrupos son consistentes con el global del estudio y no se pueden extraer más conclusiones. Aun cuando exista interacción, es necesario valorar preespecificación, plausibilidad biológica y consistencia con otros estudios, entre otros aspectos⁴.

5. Valoración del desequilibrio de ramas según presencia/ausencia de significación estadística: la significación valora la fiabilidad de un resultado entre otros estadísticamente posibles al tomar una muestra, para inferirlo a una población. Sin embargo, al comparar las características basales de las ramas, para

Tabla 1. Expresiones mejorables en informes de evaluación y posicionamiento. Aspectos estadísticos

Expresión mejorable/inaceptable	Alternativa más adecuada
1. Mención de supuestas diferencias o “tendencia” sin significación estadística: “... además, el tratamiento mostró una menor incidencia de hospitalizaciones anuales, 2,9% vs. 3,5%, si bien esta variable no llegó a alcanzar significación estadística”. O “... además, el tratamiento mostró una tendencia no significativa en la reducción de hospitalizaciones anuales”. O “... además, el tratamiento mostró una reducción no significativa en la tasa anual de hospitalización”. O “... además, el tratamiento mostró una diferencia no significativa en la tasa anual de hospitalización”.	Si no es estadísticamente significativa, no se puede expresar diferencia, superioridad ni inferioridad: “No se observó diferencia estadísticamente significativa en la tasa de hospitalización anual entre el tratamiento y el control: 2,9% vs. 3,5%; RAR -0,6% (IC95% -2,8 a 1,2%; $p=0,32$)”.
2. Uso de la reducción relativa del riesgo (RRR) para expresar el beneficio (induce sobrevaloración): “... reduce en un 33% el riesgo de fracturas no vertebrales”.	Empleo de la reducción absoluta de riesgo (RAR) (o diferencia de respuesta): “... reduce un 1% el riesgo de presentar fractura de cadera a los 3 años (2% con el tratamiento vs. 3% con placebo), con un NNT de 100 (IC95% 80 a 125): habría que tratar a 100 pacientes para, a lo largo de tres años, evitar la aparición de fractura no vertebral en uno de ellos como consecuencia del tratamiento”.
3. Uso de la RRR en el análisis de supervivencia (sobrevalora y confunde): “... el tratamiento reduce un 22% la mortalidad”.	Incluir HR y su IC95%, así como resultados de medias/medianas y su diferencia, si estas valoran adecuadamente el beneficio: “... el tratamiento mostró beneficio en supervivencia frente al comparador, con un HR de 0,78 (IC95% 0,66-0,89) y una diferencia de medianas de 2,7 (IC95% 1,0-8,1) meses (17,3 vs. 14,6 meses)”.
4. Valoración de subgrupos sin análisis de interacción, haciendo valoración intragrupal y no intergrupales: “El análisis de subgrupos presenta resultados consistentes, excepto en el marcador BRAF. Los pacientes con BRAF mutado presentan un mayor beneficio en supervivencia (HR 0,63; IC95% 0,28-0,94), mientras que los no mutados no muestran diferencias estadísticamente significativas (HR 0,91; IC95% 0,64-1,15)”.	Analizar interacción. Si es significativa o dudosa, valorar otros aspectos (preespecificación, plausibilidad biológica y consistencia). Concluir en consecuencia: “El análisis de subgrupos presenta resultados consistentes en todos los subgrupos” [Se calculó $p=0,19$]. O, si existe interacción significativa y se valoran otros aspectos: “El análisis de subgrupos presenta resultados consistentes en todos los subgrupos, excepto por posibles diferencias entre pacientes con hipertensión o no (p de interacción 0,03). Los pacientes con hipertensión presentan un resultado que podría ser mejor que el de los no hipertensos. El resultado podría explicarse por... en otro estudio similar se encontró también que... el beneficio en pacientes sin hipertensión no alcanza significación estadística (HR 0,93; IC95% 0,81-1,10) y resulta dudoso”.
5. Valoración del desequilibrio de ramas según presencia/ausencia de significación estadística: “Si bien los pacientes de alto riesgo fueron un 10% y un 18% en las ramas A y B, no hubo diferencias significativas y los grupos se consideran balanceados”.	Considerar grupos desbalanceados siempre que la magnitud y naturaleza de la diferencia sean relevantes: “Los pacientes de alto riesgo fueron un 10% y un 18% en las ramas A y B, lo que podría haber favorecido al fármaco A en la comparación”.
6. Reflejo del “ranking” como resultado de un meta-análisis en red: “El fármaco A obtuvo el 73% de probabilidades de ser el mejor tratamiento en respuesta a las 28 semanas”.	Reflejo del resultado de las comparaciones, con su intervalo de confianza/credibilidad al 95%: “Los intervalos de credibilidad no mostraron diferencias entre los fármacos de alta eficacia en respuesta a las 28 semanas. La comparación de A vs. B obtuvo una OR de 0,81 (ICr95% 0,51-1,90)”.

Tabla 2. Criterios mejorables en informes de evaluación y posicionamiento. Aspectos de redacción

Redacción mejorable	Alternativa más adecuada
<p>A. Conclusión categórica partiendo de comparaciones indirectas ajustadas: "Si bien se carece de comparaciones directas, una comparación indirecta ajustada muestra que A es superior a B, (OR 0,53; IC95% 0,41-0,73) en pacientes con recaída al cabo de un año".</p>	<p>Usar expresión que refleje la incertidumbre de este tipo de análisis: "Se carece de comparaciones directas. Con las incertidumbres de las comparaciones indirectas ajustadas, los resultados muestran que A podría ser superior a B (OR 0,53 IC95% 0, 41-0,73) en la proporción de pacientes con recaída al cabo de un año".</p>
<p>B. Valoración arriesgada de similitud o superioridad en comparaciones indirectas <i>naïve</i> (no ajustadas): "No se dispone de comparaciones directas. Pese a las limitaciones de las comparaciones indirectas, los datos muestran un efecto similar de ambas opciones de tratamiento". O, si parece que pudieran existir diferencias relevantes: "No se dispone de comparaciones directas. Pese a las limitaciones de las comparaciones indirectas, los datos parecen indicar una eficacia superior del tratamiento A frente a B".</p>	<p>Redacción prudente; mención de lo que no sabemos expresando las dudas (expresión "en negativo"): "No se dispone de comparaciones directas, ni de comparaciones indirectas ajustadas. Puesto que es metodológicamente incorrecto realizar comparaciones no ajustadas, la información disponible no permite afirmar la superioridad de un fármaco sobre otro". O bien (en situaciones específicas que lo requieran): "No se dispone de comparaciones directas, ni de comparaciones indirectas ajustadas. Puesto que es metodológicamente incorrecto realizar comparaciones indirectas no ajustadas, la información disponible no permite asegurar que el fármaco B alcance una eficacia similar a la de A".</p>
<p>C. Valoración obsoleta del comparador: "... se considera un comparador adecuado, ya que era el tratamiento de referencia en el momento en que se diseñó el ensayo"</p>	<p>Valoración actualizada del comparador: "... no es el comparador adecuado para la práctica clínica actual, ya que fue desplazado por el uso de X, que es la opción preferente en la actualidad debido a su beneficio demostrado en supervivencia global".</p>
<p>D. Dialéctica benévola o sesgada en seguridad: "Aunque el porcentaje de pacientes que presentó efectos adversos es elevado, siendo graves en el 56% de los casos, el de pacientes que suspendieron el tratamiento por efectos adversos no fue muy elevado (12% en segunda línea o superior y 22% en primera línea), por lo que se considera que el perfil de seguridad de la combinación es manejable. Además..." "La aparición de náuseas no presentó diferencias significativas entre el fármaco y el control". "La toxicidad del fármaco es superponible al placebo".</p>	<p>Referencia ecuaníme de pruebas, y siguiendo el principio de precaución: "Un 56% de los pacientes presentaron efectos adversos graves. Suspendieron el tratamiento por efectos adversos un 22% de los pacientes en primera línea y un 12% en segunda línea o superior". "Las náuseas aparecieron en un 8% de los pacientes tratados con el fármaco A y en un 2% del grupo control..." "En el estudio no apareció ningún evento adverso reseñable en comparación con el placebo".</p>
<p>E. Explicitar diferencias como "clínicamente" relevantes con variables subclínicas: "El tratamiento muestra una reducción del nivel de colesterol LDL del 53%, que se considera clínicamente relevante".</p>	<p>No concluir relevancia "clínica" con solo variables subclínicas: "El tratamiento muestra una reducción del nivel de colesterol LDL del 75%. Serán necesarios datos de morbilidad para establecer la relevancia clínica del beneficio en estos pacientes". Si la variable subclínica es la más adecuada en el escenario tratado y el beneficio es relevante, se puede concluir, p. ej.: "El tratamiento muestra beneficio significativo en SLP, con HR 0,51 (IC95% 0,32-0,74) y una diferencia de medianas de 7,3 meses (IC95% 4,3-12,2). Estos resultados se consideran relevantes para la práctica clínica" [mejor que "clínicamente relevantes", pues la variable es subclínica].</p>
<p>F. Inhibición en el posicionamiento de un fármaco por falta de evidencia: "Dada la ausencia de ensayos clínicos comparativos, en la actualidad no es posible establecer el lugar en terapéutica de este fármaco con respecto a las demás opciones en esta indicación".</p>	<p>Posicionar con los resultados de que se dispone, y reflejar la incertidumbre: "Dada la ausencia de ensayos clínicos comparativos y comparaciones indirectas ajustadas con este fármaco, con la evidencia disponible de las diferentes opciones... a) ... no se considera que en la actualidad pueda tener un lugar en terapéutica. b) ... es razonable usarlo solo cuando no sean factibles otras opciones, incluyendo... c) ... se considera como alternativa preferente el fármaco X, debido a... d) ... no es posible establecer diferencias entre ellos y se consideran opciones alternativas adecuadas/equivalentes en la misma indicación".</p>
<p>G. Posicionamiento ambiguo y trasladado al clínico por falta de evidencia: "... dada la incertidumbre antes comentada, la elección de este fármaco dependerá de una valoración individual de las características del paciente".</p>	<p>Realizar posicionamiento indicando las características diferenciales: "... dada la incertidumbre antes comentada, la elección de este fármaco tendrá en cuenta la presencia de hipertensión, historial de patología ulcerosa, presencia de leucopenia... (incluyendo detalles y valores de corte, si es posible)". Si no hay evidencia de características que dirijan la selección del tratamiento: "a) ... no se considera que en la actualidad su incorporación tenga interés en terapéutica. b) ... debería ser restringido a situaciones en que las demás opciones no puedan ser utilizadas, incluyendo... c) ... no es posible establecer diferencias entre ellos y se consideran opciones alternativas adecuadas/equivalentes en la misma indicación"</p>

valorar si un desequilibrio ha afectado a los resultados del propio estudio, no se hace inferencia alguna. Haya o no una $p < 0,05$, es preciso valorar en qué medida la magnitud y relevancia de la diferencia pueden haber influido en el resultado⁵.

6. Reflejo del "ranking" como resultado de un meta-análisis en red para la selección del tratamiento: el resultado fundamental está constituido por las propias comparaciones entre los tratamientos incluidos, con sus intervalos de confianza/credibilidad al 95%, para valorar la superioridad de un tratamiento frente a otro con un error alfa del 5%. El ranking resulta estadísticamente poco sólido para establecer conclusiones de superioridad⁶.

Aspectos de criterio y redacción

A. Valoración categórica de resultados en comparaciones indirectas ajustadas: debemos evitar expresar los resultados de las comparaciones indirectas ajustadas con el mismo nivel de certeza que el empleado para un ensayo clínico, ya que están sujetas a mayor incertidumbre⁷.

B. Valoración arriesgada de similitud o superioridad en comparaciones indirectas *naïve* (no ajustadas): la valoración de resultados en ensayos clínicos diferentes sin una comparación indirecta ajustada podría ser un recurso extremo, cuando no sea posible otra aproximación. Tal comparación no puede sustentar ninguna afirmación, ni fundamentar las conclusiones del informe⁸.

C. Valoración obsoleta del comparador: la evaluación no juzga la decisión del investigador al elegir, en su día, el comparador, sino que precisa valorar el beneficio diferencial para el paciente frente a los tratamientos actuales. Se debe valorar en qué medida el comparador utilizado permite extraer conclusiones para la terapia actual.

D. Dialéctica sesgada en seguridad: la exposición debe ser lo más ecuánime posible y evitar expresiones generalizadoras y ambiguas como, p.ej., "perfil de seguridad manejable/aceptable". Es mejor reflejar los resultados. Con la excepción de ensayos con objetivo principal de seguridad, los estudios no se diseñan con potencia estadística suficiente para detectar diferencias significativas en toxicidad. Por tanto, se sigue el principio de precaución, considerando aquellos efectos adversos con un posible aumento de incidencia y que podrían ser atribuibles al fármaco. Ningún fármaco eficaz resulta "superponible al placebo" en seguridad.

E. Valorar diferencias como "clínicamente" relevantes con variables subclínicas: expresar que un resultado es clínicamente relevante exige disponer de una variable "clínica" u orientada al paciente⁹ (no una variable subclínica, como colesterolemia, tensión arterial, supervivencia libre de progresión (SLP), aumento de plaquetas...), y en la que se haya observado una diferencia que considere "relevante". Cuando la variable subclínica resulta óptima para valorar el tratamiento en la práctica –por ejemplo, la supervivencia libre de enfermedad en adyuvancia en cáncer de mama precoz, o la carga viral indetectable en VIH–, se puede expresar mejor como: "la diferencia se considera relevante para la práctica clínica".

F. Inhibición en el posicionamiento por falta de evidencia sobre un fármaco comercializado: el posicionamiento no es un ejercicio académico, sino una toma de decisiones basada en evidencia que a menudo puede ser precaria. Debe evitarse la inhibición, a menos que se piense que el comité que elabora el informe no es adecuado para decidir y que debería hacerlo otro estamento o el prescriptor individual (a un nivel u otro, se hará con la información disponible). Las diferentes conclusiones dependerán de la evidencia con las otras opciones, los

resultados del fármaco y su fiabilidad, y de la utilidad de incorporar el nuevo fármaco al arsenal terapéutico.

G. Posicionamiento ambiguo y trasladado al clínico: todos los fármacos son prescritos tras una "valoración individual de las características del paciente", por lo que señalar esto como posicionamiento de forma genérica sería una obviedad. Si hay características que, según la evidencia disponible, deban valorarse para la selección de uno u otro tratamiento, deben enunciarse y justificarse. Si no las hay, el prescriptor no va a disponer de más criterios que los que ya habitualmente se aplican en la práctica clínica, y eso no obsta para tomar una decisión de posicionamiento.

Finalmente, conviene aclarar un par de conceptos utilizados en posicionamiento:

"Preferente/s": fármaco/s considerados como primera opción aconsejable para la situación evaluada. Es distinto de "primera línea". Pueden existir fármacos preferentes en primera línea, en segunda línea, etc.

"Alternativas terapéuticas similares/equivalentes": fármacos para los que, con la evidencia disponible, no es posible establecer preferencia de unos sobre otros en la mayoría de los pacientes en una determinada situación clínica y que, por tanto, resulta razonable seleccionarlos según costes, sin perjuicio de que existan pacientes en los que sí se pueda considerar una preferencia justificada para la selección del tratamiento¹⁰.

La síntesis de estos aspectos puede tener una utilidad docente y de apoyo específico a las labores de redacción/visión de los informes de evaluación y posicionamiento de nuevos medicamentos. Muchos de ellos pueden ser útiles también para la redacción o revisión de otros trabajos científicos.

Agradecimientos: A Esther Márquez Saavedra, Jefa de Servicio de Farmacia Sevilla Sur, por su revisión final del manuscrito y sus valiosas aportaciones al mismo.

Conflicto de intereses: Los autores declaran no tener conflicto de intereses.

BIBLIOGRAFÍA

- Gibbs NM, Gibbs SV. Misuse of 'trend' to describe 'almost significant' differences in anaesthesia research. *Br J Anaesth*. 2015;115(3):337-9.
- Guthrie B, Thompson A, Dumbreck S, et al. Comparing treatments in terms of absolute benefit [Internet]. NIHR JournalsLibrary; 2017 Disponible en: <https://www.ncbi.nlm.nih.gov/books/NBK426103/> [consultada 02/12/2020].
- Sun X, Ioannidis JPA, Agoritsas T, et al. How to use a subgroup analysis: users' guide to the medical literature. *JAMA*. 2014;311(4):405-11.
- Gil-Sierra MD, Fénix-Caballero S, Abdelkader-Martin L, et al. Checklist for clinical applicability of subgroup analysis. *J Clin Pharm Ther*. 2020;45(3):530-8.
- Roberts C, Torgerson DJ. Understanding controlled trials: Baseline imbalance in randomised controlled trials. *Br Med J*. 1999;319(7203):185.
- Mbuagbaw L, Rochweg B, Jaeschke R, et al. Approaches to interpreting and choosing the best treatments in network meta-analyses. *Syst Rev*. 2017;6(1):79.
- Ortega A, Fraga MD, Alegre-del-Rey EJ, Puigventós-Latorre F, et al. A checklist for critical appraisal of indirect comparisons. *Int J Clin Pract*. 2014;68(10):1181-9.
- Ortega Eslava A, Fraga Fuentes MD, Alegre Del Rey EJ, et al. Comparaciones indirectas. *Farm Hosp*. 2012;36(4):173-5.
- Ebell MH, Siwek J, Weiss BD, et al. Strength of Recommendation Taxonomy (SORT): A Patient-Centered Approach to Grading Evidence in the Medical Literature. *Am Fam Physician*. 2004;69(3):548-56.
- Alegre del Rey EJ, Fénix Caballero S, Castaño Lara R, et al. Evaluación y posicionamiento de medicamentos como alternativas terapéuticas equivalentes. *Med Clin (Barc)*. 2014;143(2):85-90.

