



TESIS DOCTORAL

PROGRAMA DE DOCTORADO EN FARMACIA

**Desarrollo de la aplicabilidad clínica y
farmacoeconómica del análisis por subgrupos
según la medicina basada en la evidencia.**

Manuel David Gil Sierra

Sevilla, 2023



DEPARTAMENTO DE FARMACOLOGÍA

FACULTAD DE FARMACIA

UNIVERSIDAD DE SEVILLA

La Dra. Marina Sánchez Hidalgo, Profesora Titular de Universidad, y la Dra. Catalina Alarcón de la Lastra Romero, Catedrática de Universidad, adscritas al Departamento de Farmacología de la Facultad de Farmacia de la Universidad de Sevilla, como directoras de la Tesis Doctoral titulada “Desarrollo de la aplicabilidad clínica y farmacoeconómica del análisis por subgrupos según la medicina basada en la evidencia”, realizada por Manuel David Gil Sierra para aspirar al grado de Doctor en Farmacia, certifican que el presente trabajo ha sido realizado por el doctorando bajo nuestra dirección y corresponde fielmente a los resultados obtenidos.

Y para que conste, firmo el presente, en Sevilla a 25 de enero de 2023.

Fdo.: Dra. Marina Sánchez Hidalgo

Fdo.: Dra. Catalina Alarcón de la Lastra Romero

PUBLICACIONES, GALARDONES Y PONENCIAS

Los resultados incluidos en la presente Tesis Doctoral han sido recogidos en los siguientes **artículos científicos**:

1. Gil-Sierra MD, Fénix-Caballero S, Abdel Kader-Martin L, Fraga-Fuentes MD, Sánchez-Hidalgo M, Alarcón de la Lastra-Romero C, et al. Checklist for clinical applicability of subgroup analysis. *J Clin Pharm Ther.* 2020;45(3):530-8. Doi: 10.1111/jcpt.13102
2. Gil-Sierra MD, Briceño-Casado MDP, Fénix-Caballero S, Alegre-Del Rey EJ, Alarcón de la Lastra-Romero C, Sánchez-Hidalgo M. Daratumumab-based therapies in transplant-ineligible patients with untreated multiple myeloma and hepatic dysfunction: A systematic review of subgroup analyses. *J Oncol Pharm Pract.* 2023;29(1):155-161. Doi: 10.1177/10781552211062144
3. Gil-Sierra MD, Alegre-Del Rey EJ, Alarcon de la Lastra-Romero C, Sánchez-Hidalgo M. Use of docetaxel in low- and high-burden metastatic hormone-sensitive prostate cancer: A systematic review and assessment of subgroup analyses. *J Oncol Pharm Pract.* 2021;27(7):1743-1750. Doi: 10.1177/10781552211037322
4. Gil-Sierra MD, Briceño-Casado MP, Alegre-Del Rey EJ, Sánchez-Hidalgo M. Efficacy of early use of remdesivir: a systematic review of subgroup analysis. *Rev Esp Quimioter.* 2022;35(3):249-259. Doi: 10.37201/req/154.2021
5. Gil-Sierra MD, Fenix-Caballero S, Alegre-del Rey E. Pembrolizumab plus chemotherapy in lung cancer. *N Engl J Med.* 2018;379(11):e18. Doi: 10.1056/NEJMc1808567
6. Gil-Sierra MD, Briceño-Casado MDP, Fénix-Caballero S. Analysis of overall survival benefit of abemaciclib plus fulvestrant in hormone receptor-positive, ERBB2-negative breast cancer. *JAMA Oncol.* 2020;6(7):1122. Doi: 10.1001/jamaoncol.2020.1516
7. Gil-Sierra MD, Sánchez-Hidalgo M, Alarcón de la Lastra-Romero C. Clinical decision-making about neoadjuvant nivolumab plus ipilimumab. *JAMA Oncol.* 2021;7(2):309. Doi: 10.1001/jamaoncol.2020.6989

8. Gil-Sierra MD, Briceño-Casado MDP, Sánchez-Hidalgo M. Nivolumab plus ipilimumab for previously treated stage IV squamous cell lung cancer. *JAMA Oncol.* 2022;8(4):642. Doi: 10.1001/jamaoncol.2021.7787
9. Alegre-Del Rey EJ, Gil-Sierra MD, Alarcón de la Lastra-Romero C, Sánchez-Hidalgo M. Remdesivir and mortality reduction in COVID-19 patients: a systematized subgroup analysis of clinical trials. *Farm Hosp.* 2021;45(1):28-31. Doi: 10.7399/fh.11591.
10. García-Mochón L, Gil-Sierra MD, Alegre-Del Rey EJ, Alarcón de la Lastra-Romero C, Sánchez-Hidalgo M. Economic evaluation and budgetary burden of mepolizumab in severe refractory eosinophilic asthma. *Farm Hosp.* 2019;43(6):187-93. Doi: 10.7399/fh.11221.
11. Gil-Sierra MD, Briceño-Casado MDP, Sánchez-Hidalgo M. Clinical and pharmacoeconomic impact of subgroup analysis in onco-hematological patients. *Support Care Cancer.* 2022;30(5):3761-72. Doi: 10.1007/s00520-022-06823-y

Asimismo, se presentaron las siguientes **comunicaciones a congresos**:

1. Gil Sierra MD, Fénix Caballero S, Abdel Kader Martín L, Fraga Fuentes MD, Sánchez Hidalgo M, Alarcón de la Lastra Romero C, et al. Cuestionario para valorar la aplicabilidad del análisis por subgrupos a la toma de decisiones terapéuticas. 63º Congreso Nacional de la Sociedad Española de Farmacia Hospitalaria (SEFH). Palma de Mallorca, del 8 al 10 de noviembre de 2018.
2. Gil Sierra MD, Primo Vera J, Alegre del Rey EJ. Calculadora de test de interacción para el análisis de subgrupos. XVI Congreso de la Sociedad Andaluza de Farmacéuticos de Hospitales y Centros Sociosanitarios (SAFH). El Rompido (Huelva), del 24 al 26 de abril de 2019.
3. Gil-Sierra MD, Fénix-Caballero S, Briceño-Casado MDP, Sánchez-Hidalgo M, Alarcón de la Lastra-Romero C, De la Calle-Riaguas B, et al. Use of daratumumab based treatments in patients with multiple myeloma and hepatic impairment. European Association of Hospital Pharmacists (EAHP) 25th Anniversary Congress. Virtual, del 23 al 28 marzo de 2021.

4. Gil-Sierra MD, Briceño-Casado MDP, Sánchez-Hidalgo M, Alarcón de la Lastra-Romero C, De la Calle-Riaguas B, Dominguez-Cantero M, et al. Evaluation of docetaxel in low and high burden metastatic hormone sensitive prostate cancer. European Association of Hospital Pharmacists (EAHP) 25th Anniversary Congress. Virtual, del 23 al 28 marzo de 2021.
5. Gil-Sierra MD, Briceño-Casado MDP, Sánchez-Hidalgo M, Alarcón de la Lastra-Romero C, De la Calle-Riaguas B, Dominguez-Cantero M, et al. Subgroup analysis about efficacy of early use of remdesivir in COVID-19. European Association of Hospital Pharmacists (EAHP) 25th Anniversary Congress. Virtual, del 23 al 28 marzo de 2021.
6. Gil-Sierra MD, Fenix-Caballero S, Sanchez-Hidalgo M, Alarcon de la Lastra Romero C, Briceno-Casado MDP, Rios-Sanchez E, et al. Indirect comparison of pembrolizumab plus chemotherapy versus pembrolizumab in lung cancer. European Association of Hospital Pharmacists (EAHP) 24th Anniversary Congress. Barcelona, del 27 al 28 de marzo 2019.
7. Gil Sierra MD, García Mochón L, Alegre del Rey EJ, Borrero Rubio JM, Alarcón de la Lastra Romero C, Sánchez Hidalgo M. Aplicación farmacoeconómica del análisis por subgrupos en asma eosinofílica refractaria grave. XV Congreso de la Sociedad Andaluza de Farmacéuticos de Hospitales y Centros Sociosanitarios (SAFH). Almería, del 11 al 13 abril 2018.
8. Gil-Sierra MD, Gimeno-Ballester V, Briceno-Casado MDP, Sanchez-Hidalgo M, Alarcon de la Lastra-Romero C, Fenix-Caballero S, et al. Network meta-analysis of first-line antiangiogenic drugs in advanced renal cell carcinoma. European Association of Hospital Pharmacists (EAHP) 24th Anniversary Congress. Barcelona, del 27 al 28 de marzo 2019.

De estas comunicaciones a congresos, se recibieron las siguientes **nominaciones a premio**:

1. Gil Sierra MD, Fénix Caballero S, Abdel Kader Martín L, Fraga Fuentes MD, Sánchez Hidalgo M, Alarcón de la Lastra Romero C, et al. Cuestionario para valorar la aplicabilidad del análisis por subgrupos a la toma de decisiones terapéuticas. 63^º Congreso Nacional de la Sociedad Española de Farmacia Hospitalaria (SEFH). Palma de Mallorca, del 8 al 10 de noviembre de 2018.
2. Gil Sierra MD, Primo Vera J, Alegre del Rey EJ. Calculadora de test de interacción para el análisis de subgrupos. XVI Congreso de la Sociedad Andaluza de Farmacéuticos de

Hospitales y Centros Sociosanitarios (SAFH). El Rompido (Huelva), del 24 al 26 de abril de 2019.

3. Gil-Sierra MD, Gimeno-Ballester V, Briceno-Casado MDP, Sanchez-Hidalgo M, Alarcon de la Lastra-Romero C, Fenix-Caballero S, et al. Network meta-analysis of first-line antiangiogenic drugs in advanced renal cell carcinoma. European Association of Hospital Pharmacists (EAHP) 24th Anniversary Congress. Barcelona, del 27 al 28 de marzo 2019.

Este trabajo ha obtenido los siguientes **premios**:

1. Premio de la Sociedad Española de Farmacia Hospitalaria a la Mejor Comunicación Oral Científica del 63^º Congreso de la Sociedad Española de Farmacia Hospitalaria, en Palma de Mallorca, del 8 al 10 de noviembre de 2018, titulada “Cuestionario para valorar la aplicabilidad del análisis por subgrupos a la toma de decisiones terapéuticas” y cuyos autores son Gil Sierra MD, Fénix Caballero S, Abdel Kader Martín L, Fraga Fuentes MD, Sánchez Hidalgo M, Alarcón de la Lastra Romero C, et al.

2. Premio J.R Ferrándiz (primer áccesit) de la Fundación Española de Farmacia Hospitalaria (año 2021) a la publicación “Remdesivir y reducción de mortalidad en pacientes con COVID-19: análisis sistematizado de subgrupos de los ensayos clínicos” cuyos autores son Alegre-Del Rey EJ, Gil-Sierra MD, Alarcón de la Lastra-Romero C, Sánchez-Hidalgo M.

Se realizaron las siguientes **ponencias**:

1. “Colaboración de farmacéuticos con economistas para realizar estudios farmacoeconómicos” en el XVI Congreso de la Sociedad Andaluza de Farmacéuticos de Hospitales y Centros Sociosanitarios (SAFH), celebrado en el Rompido (Huelva) del 24 al 26 de abril de 2019. Actividad Acreditada por la Secretaría General de Investigación, Desarrollo e Innovación en Salud de la Consejería de Salud y Familias de la Junta de Andalucía.

2. “El análisis por subgrupos en la evidencia científica. Una herramienta sanitaria transversal” en las Sesiones Clínicas de Actualización para Facultativos del Hospital Nuestra Señora de Guadalupe del Servicio Canario de Salud, celebrada el 28 de noviembre de 2019.

Actividad Acreditada por la Comisión Canaria de Formación Continuada de las Profesiones Sanitarias

3. “La aplicabilidad del análisis por subgrupos en onco-hematología” en la Reunión Virtual de la Zona de Canarias de la Sociedad Española de Farmacia Hospitalaria (SEFH), celebrada el 12 junio 2020.
4. “Farmacéutico hospitalario en la optimización farmacoeconómica del asma” en la Jornada Farmacoeconómica de Canarias del Servicio Canario de Salud, celebrada el 16 marzo 2021.
5. “Comparaciones indirectas y análisis por subgrupos para posicionar fármacos: ejemplo ilustrativo” en el Webinar 13 de La FHactoria, celebrada el 24 de mayo de 2021.
6. “Módulo de tecnologías: Análisis de datos” del Proyecto Cápsulas de Innovación ViiV, celebrada el 10 de mayo 2022.
7. “Impacto clínico y farmacoeconómico de la aplicabilidad del análisis por subgrupos” como MasterClass en La FHactoria, celebrada el 1 de septiembre de 2022.

AGRADECIMIENTOS

Quisiera comenzar por agradecer a todas aquellas personas que he conocido en mi vida que se esfuerzan por promover el desarrollo profesional de los demás, sea cual sea el ámbito, aunque especialmente el educativo y sanitario. En este caso habría que destacar a Marina Sánchez y Catalina Alarcón de la Lastra por colaborar conmigo en este arduo y largo camino. También me gustaría agradecer a Emilio Jesús Alegre del Rey por haberme ayudado a dar los primeros pasos de esta Tesis Doctoral.

Asimismo, como voy a poder olvidar a aquellos familiares y amigos que me han apoyado y dado aliento en los peores momentos. Quiero acordarme especialmente de mi pareja, que ha estado soportando durante años las consecuencias de la dura conciliación académica, profesional y personal que sufrimos los sanitarios, sobre todo en tiempos de pandemia. Nunca habríamos imaginado que sería tan difícil, pero todo lo bueno cuesta un poco más.

Por último, y por encima de todas las cosas, quisiera agradecer a mis padres no sólo el apoyo que me han dado para poder empezar (y acabar) esta Tesis Doctoral, sino durante toda mi vida. Los valores que me han inculcado desde pequeño: la honestidad, la determinación y el esfuerzo. No tuvieron que darme lecciones de nada, sólo bastaba con mirar sus actos. Nunca negociaron el esfuerzo y gracias a ese tesón que desprenden pude llegar más lejos de lo que jamás hubiera imaginado. Durante mis ya más de 30 años han sido el verdadero motor moral de mis avances, y espero que lo sigan siendo durante mucho tiempo más. Esto va por y para ellos.

ÍNDICE

ABREVIATURAS	12
RESUMEN.....	14
1. INTRODUCCIÓN	17
2. JUSTIFICACIÓN Y OBJETIVOS.....	31
3. RESULTADOS	36
3.1. Desarrollo de una herramienta para la interpretación y aplicación del análisis por subgrupos.....	38
3.2. Revisión sistemática de la bibliografía disponible sobre análisis por subgrupos de contextos clínicos de alto impacto.	55
3.2.1. Esquemas con daratumumab en pacientes con mieloma múltiple de nuevo diagnóstico no candidatos a trasplante que presentan insuficiencia hepática.....	58
3.2.2. Docetaxel en cáncer de próstata metastásico hormonosensible con alta y baja carga tumoral.....	69
3.2.3. Uso temprano de remdesivir en pacientes con COVID-19.	83
3.3. Aplicación de la metodología y conocimientos desarrollados sobre publicaciones concretas.....	105
3.3.1. Pembrolizumab más quimioterapia en cáncer de pulmón no microcítico (CPNM) metastásico.....	108
3.3.2. Abemaciclib más fulvestrant en cáncer de mama con receptor hormonal positivo y ERBB2 negativo tras progresión a terapia endocrina.....	115
3.3.3. Nivolumab más ipilimumab en carcinoma escamoso de la cavidad oral no tratado previamente	122
3.3.4. Nivolumab más ipilimumab en pacientes con CPNM escamoso estadio IV tratado previamente	128
3.3.5. Remdesivir y reducción de mortalidad en pacientes con COVID-19.	135
3.4. Estudio fármaco-económico sobre subgrupos de mepolizumab en asma eosinofílica refractaria grave	145
4. DISCUSIÓN GENERAL.....	162
5. CONCLUSIONES.....	180
6. BIBLIOGRAFÍA.....	184

ABREVIATURAS

ADT = terapia de deprivación androgénica, del inglés *androgen deprivation therapy*

AIP = análisis de impacto presupuestario

ARN = ácido ribonucleico

CD4 = cúmulo de diferenciación 4, del inglés *cluster of quadruple differentiation*

COVID-19 = enfermedad por coronavirus 2019

CPNM = cáncer de pulmón no microcítico

CSI = corticoides sistémicos inhalados

CSO = corticoides sistémicos orales

ECA = ensayo clínico aleatorizado

ECOG = Eastern Cooperative Oncology Group

et al. = y colaboradores

GÉNESIS = Grupo de Evaluación de Novedades, Estandarización e Investigación en Selección de Medicamentos

GHEMA = Grupo Hospitalario de Evaluación de Medicamentos de Andalucía

HR = *hazard ratio*

IC = intervalo de confianza

IgE = inmunoglobulina E

IPT = Informes de Posicionamiento Terapéutico

ISS = sistema internacional de estadificación, del inglés *International Staging System*

LABA = agonistas β 2 de larga duración, del inglés *long-acting beta-adrenoceptor agonist*

Ln = logaritmo neperiano

MADRE = Método de Ayuda para la toma de Decisiones y la Realización de Evaluaciones de medicamentos

MEPO: mepolizumab.

mHSPC = cáncer de próstata metastásico hormonosensible

MM = mieloma múltiple

Mt/Mb = mutaciones por megabase

N+I = combinación de nivolumab más ipilimumab

NAF = no de alto flujo

NMA = network meta-analysis

NNT = número necesario de pacientes a tratar, del inglés *number needed to treat*

OMA: omalizumab

OMS = Organización Mundial de la Salud

Pb-CT = pembrolizumab mas quimioterapia

PD-L1 = ligando de muerte programada 1, del inglés *programmed death ligand 1*

PICOS = modelo Población-Intervención-Comparador-Resultado-Estudio

PRISMA = guía Preferred Reporting Items for a Systematic Review and Meta-analysis

PSA = antígeno prostático específico, del inglés *prostate-specific antigen*

RdR = razón de tasas de mortalidad, del inglés *ratio of death rates*

RR = riesgo relativo

RRR = reducción relativa del riesgo

SARS-CoV-2 = síndrome de infección respiratoria severa aguda del coronavirus 2

SG = supervivencia global

SLP = supervivencia libre de progresión

SNS = Sistema Nacional de Salud

TMB = carga de mutación tumoral, del inglés *tumor mutation burden*

RESUMEN

El análisis por subgrupos evalúa los efectos de fármacos en poblaciones según una o varias características concretas. En la actualidad, no existe un consenso homogéneo sobre el manejo de los resultados por subgrupos. Una inadecuada interpretación y aplicación del análisis por subgrupos puede tener impacto negativo tanto a nivel individual (paciente concreto) como colectivo (sistema sanitario). Los objetivos de esta Tesis Doctoral fueron los siguientes: (1) elaborar una metodología sistemática para el abordaje del análisis por subgrupos; (2) aplicar esta nueva herramienta en contextos clínicos de elevado impacto clínico, fármaco-económico y social; y (3) desarrollar un estudio fármaco-económico sobre mepolizumab en el tratamiento del asma eosinofílica refractaria grave considerando los datos de análisis por subgrupos.

Primero, se diseñó y validó una herramienta para valorar la aplicabilidad del análisis por subgrupos. Esta herramienta está compuesta por unas cuestiones preliminares para descartar análisis por subgrupos sin condiciones mínimas relevantes, y un cuestionario. Una respuesta negativa en alguna de las cuestiones preliminares descarta la aplicabilidad del análisis por subgrupos de forma directa. Por otra parte, el cuestionario asocia valoraciones con puntuaciones para cada criterio de interpretación del análisis por subgrupos: asociación estadística, que incluye $p(i)$, pre-especificación de subgrupos, tamaño muestral, número de factores estudiados y resultado global del estudio; plausibilidad biológica y consistencia de los resultados. La herramienta sistemática obtuvo valoraciones positivas por los usuarios, mostrándose útil y fácil de manejar, por lo que se presenta como una metodología con la que diferentes profesionales sanitarios pueden entablar discusiones objetivas en la toma de decisiones terapéuticas que involucran a análisis por subgrupos.

Posteriormente, se aplicó la herramienta desarrollada a los resultados de revisiones sistemáticas de la bibliografía disponible sobre análisis por subgrupos en contextos clínicos de alto impacto. Los fármacos abordados fueron daratumumab en mieloma múltiple,

docetaxel en cáncer de próstata y remdesivir en COVID-19. Además, se aplicó la metodología validada para la interpretación de análisis por subgrupos en publicaciones concretas de ensayos clínicos que realizan afirmaciones inapropiadas. Los fármacos involucrados en esta parte fueron pembrolizumab en cáncer de pulmón, abemaciclib en cáncer de mama, la combinación de nivolumab más ipilimumab en varios tipos de cáncer, y remdesivir en COVID-19. La herramienta demostró que puede evitar un impacto negativo de afirmaciones erróneas sobre los resultados de subgrupos en la eficacia, seguridad, adecuación y posicionamiento terapéutico de los fármacos, además de minimizar el sesgo en la selección de datos para la realización de nuevos estudios.

Para la consecución del tercer objetivo se desarrolló un estudio fármaco-económico sobre el uso de mepolizumab en asma eosinofílica refractaria grave que favoreció la priorización de mepolizumab en pacientes con asma eosinofílica refractaria grave no mediada por IgE y niveles plasmáticos ≥ 500 eosinófilos/ μL , que es el subgrupo en el que se observó una mayor eficiencia. Por tanto, esto contribuye a la sostenibilidad del sistema sanitario. En conclusión, esta Tesis Doctoral ha permitido desarrollar y validar una metodología sistemática y ordenada sobre análisis por subgrupos que tiene en cuenta criterios objetivos para favorecer una toma de decisiones clínicas racional, y así evitar importantes repercusiones negativas en los pacientes o mejorando la gestión de los sistemas de salud, tal y como hemos demostrado al aplicarla a numerosos casos de elevado impacto clínico y fármaco-económico.

1. INTRODUCCIÓN

1. INTRODUCCIÓN

El análisis por subgrupos evalúa el efecto de una intervención terapéutica –frecuentemente de fármacos- en poblaciones según una o varias características concretas (1,2). En la actualidad, estas determinaciones suelen llevarse a cabo con carácter exploratorio para generar nuevas hipótesis sobre el perfil de los pacientes que se benefician en mayor o menor medida. De esta manera, el análisis por subgrupos ha alcanzado un considerable impacto en las conclusiones de los ensayos clínicos aleatorizados (ECA) (3). Estos resultados podrían utilizarse para desechar el uso de tratamientos de dudoso beneficio, y así evitar riesgos innecesarios de seguridad o suponer una interesante contribución en la mejora de la eficiencia del sistema sanitario público. Por otro lado, podría detectar subpoblaciones con beneficio relevante cuando el resultado en la población global no lo presenta.

El carácter exploratorio del análisis de subgrupos en los ensayos clínicos requiere que los hallazgos observados sean contrastados en un nuevo ensayo clínico para asegurar los resultados con fiabilidad estadística. Generalmente, estos estudios confirmatorios nunca se realizan y las decisiones en la práctica clínica tienen que tomarse con esa carencia de información. Las suposiciones sugeridas en estas investigaciones provenientes del análisis por subgrupos son confirmadas en muy pocos casos (4). Se podría desechar por completo los análisis por subgrupos. No obstante, obviar estos datos es también una decisión que tiene sus consecuencias, y resulta especialmente controvertida en situaciones con riesgos muy elevados o costes difícilmente asumibles, las cuales no son infrecuentes en la práctica.

La interpretación del análisis por subgrupos debe ser desarrollada con cautela, ya que a aplicación del análisis de subgrupos conlleva un grado importante de incertidumbre. Esto se debe al aumento de los errores α y β (5). El error de tipo α (también denominado falso positivo) es el que se comete cuando se rechaza una hipótesis nula que, en realidad, es verdadera. Esto ocurre cuando se asume unas diferencias entre las hipótesis que en realidad son inexistentes. El error de tipo β (también denominado falso negativo) es aquel que se produce cuando no se rechaza la hipótesis nula que, realmente, es falsa. Esto sucede al concluir que no existen diferencias que realmente existen. En el análisis por subgrupos,

la realización de determinaciones adicionales y la comparación de nuevas redistribuciones de pacientes en cada una de las ramas –una nueva para cada característica que divide la población en subgrupos–, aumenta la posibilidad de obtener diferencias con un determinado nivel de significación. Estas diferencias pueden deberse al azar o por una distribución desbalanceada de factores relacionados con el beneficio. Por tanto, se podrían generar diferencias aparentes que no son reales y el error α se ve incrementado. Por otra parte, el error β aumenta cuando se produce la división en subgrupos de una población con tamaño muestral mínimo necesario para alcanzar un poder estadístico suficiente en un ensayo clínico, provocando la incapacidad de detectar diferencias existentes entre subgrupos.

Análisis de datos

Los resultados de subgrupos pueden ser representados de varias maneras en los artículos científicos. Por una parte, se pueden describir los resultados obtenidos en los diferentes subgrupos en el texto de las publicaciones (sin gráficas) o en tablas. Por otra parte, se pueden representar los datos de los subgrupos mediante gráficas. Un ejemplo pueden ser las figuras de las curvas de supervivencia de las subpoblaciones en los procesos neoplásicos. Estas representaciones suelen dar lugar a interpretaciones inadecuadas de los análisis por subgrupos, ya que diferencias aparentes pueden ser asumidas sin que exista un respaldo estadístico sólido.

No obstante, la mayoría de los estudios representan los análisis por subgrupos de eficacia de los tratamientos a través de *forest plot* (6,7) (**Figura 1**). De esta manera, se muestran los resultados de los subgrupos de forma conjunta y complementaria. Se describen los factores que dividen a la población, como puede ser el género o la presencia de una mutación, y en función de cada uno de estos factores, se disponen los datos mediante un valor central y un intervalo de confianza (IC) para cada subgrupo. Esto permite comprobar el solapamiento de los ICs. Esta forma de mostrar los datos también permite ver el beneficio de la intervención sanitaria evaluada frente al comparador usando una línea vertical. Aunque

esta representación pudiera parecer sencilla, se han descrito importantes errores de interpretación del análisis por subgrupos en la literatura (8).

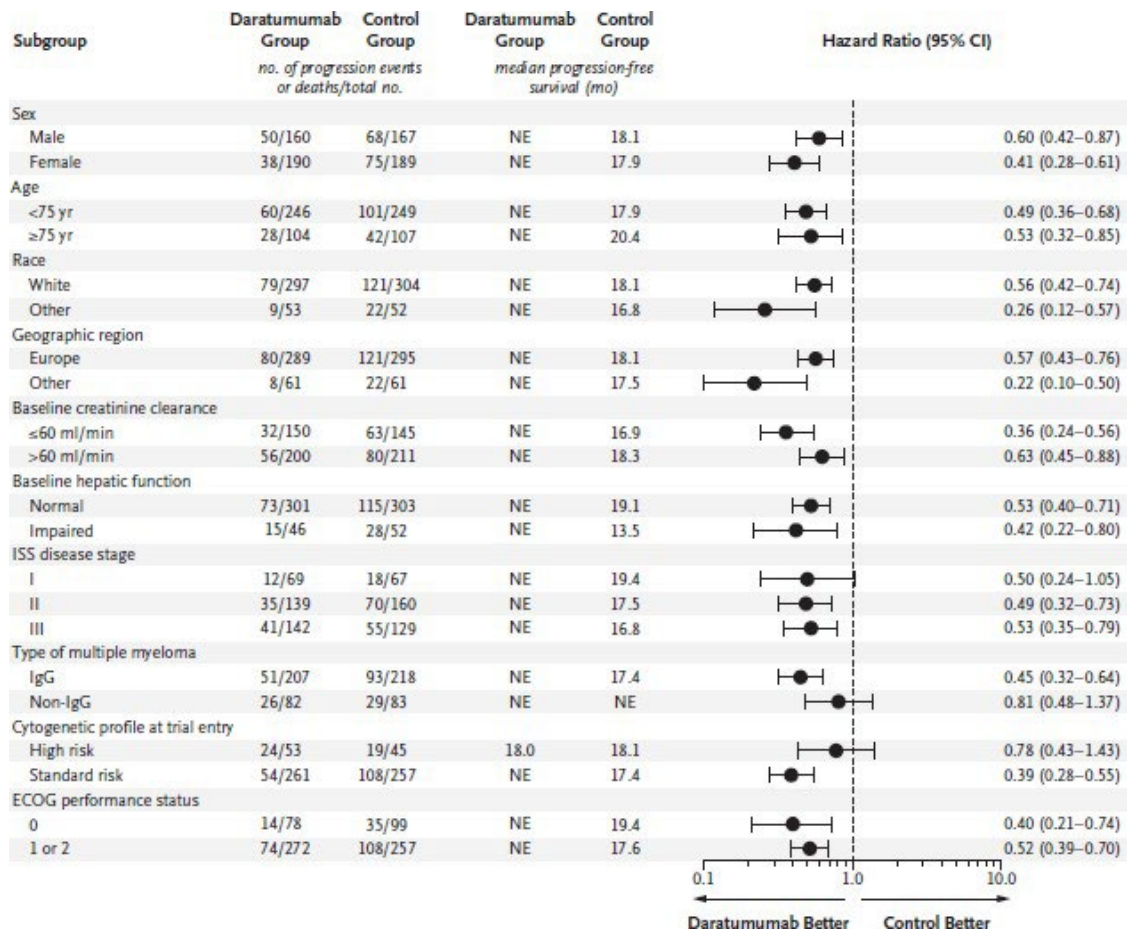


Figura 1. Ejemplo de representación de análisis por subgrupos. Tomada de Mateos *et al.* (2018) (7).

Tipos de errores de interpretación

Existe un amplio abanico de errores de interpretación de los datos de subgrupos. Para hacer una correcta valoración de estos resultados, es indispensable entender diversos conceptos, como la necesidad de disponer de subgrupos complementarios o el valor neutro, el cual representa a la ausencia de diferencia entre la intervención evaluada y el grupo control. Para poder realizar conclusiones acertadas sobre la influencia de una característica de la población en el efecto de un tratamiento, debemos poder consultar

tanto los datos de un subgrupo como los del complementario. Por ejemplo, para saber si una determinada terapia alcanza mejores respuestas en los pacientes mayores de 65 años, tenemos que disponer tanto de los resultados obtenidos en la población menor como en la mayor de 65 años. Asimismo, la línea vertical discontinua que se ha mostrado anteriormente en los *forest plot* se coloca encima del valor neutro, marcando el límite entre la ventaja de uso de la intervención o del brazo control. Las determinaciones que estén posicionadas en el lado del grupo con la intervención representan un efecto favorable de usar dicha intervención, mientras que los valores que se encuentren desplazados hacia la parte del brazo control indican una ventaja numérica del control usado en el estudio.

Un error frecuente es la consideración de una mejor eficacia de un fármaco en un subgrupo frente al efecto en la población complementaria cuando el IC del primer subgrupo no atraviesa el valor neutro, mientras que el segundo subgrupo tiene un IC a través del valor neutro (4,9,10). Este hallazgo no tiene relevancia por sí sólo, ya que puede ser consecuencia de del menor tamaño muestral de las subpoblaciones y redistribución de los factores relacionados con el resultado en los pacientes. Un ejemplo gráfico puede observarse en la **Figura 1**, donde el tratamiento con daratumumab no muestra diferencias en los subgrupos de los estadios de enfermedad según *International Staging System (ISS)*, pero el estadio I atraviesa el valor neutro (7). Este hecho podría dar lugar a afirmaciones erróneas sobre un menor efecto del esquema basado en daratumumab en esta subpoblación. Por tanto, la consideración de distintos comportamientos entre subgrupos sin diferencias estadísticamente significativas podría tener en cuenta desviaciones numéricas compatibles con el azar en la toma de decisiones clínicas.

Otro error que se puede cometer es la aceptación de diferencias de efecto valorando exclusivamente la posición del valor central de los ICs de los subgrupos. Por ejemplo, puede surgir la idea de usar sólo el tratamiento intervención en hombres cuando el valor central del IC del subgrupo de varones se encuentra en el lado de beneficio por el tratamiento intervención (con respecto a la línea vertical que representa al valor neutro), y el valor central en el subgrupo de mujeres se encuentra en el lado de beneficio por el tratamiento

control. Sin embargo, la interpretación del análisis por subgrupos debe atender a otros criterios que serán detallados más adelante.

Asimismo, aquellos estudios que muestran un resultado global sin diferencias estadísticamente significativas entre el brazo con intervención y el grupo control son susceptibles de tener interpretaciones inadecuadas el análisis por subgrupos (11), ya que si en estos casos se observan diferencias estadísticamente significativas entre subgrupos debe descartarse que este hallazgo sea un “artefacto” que se deba al azar, pues el objetivo principal del ensayo clínico –determinar diferencias entre el grupo intervención y el grupo control- no fue alcanzado. Por tanto, es imprescindible considerar adecuada y prudentemente el análisis por subgrupos en la toma de decisiones en la práctica clínica, usando criterios metodológicos estrictos (4).

Criterios a valorar

Para atenuar las limitaciones del análisis por subgrupos, se pueden tener en cuenta una serie de criterios fundamentales: interacción estadística, preespecificación de los subgrupos, plausibilidad biológica y consistencia con otros estudios

- Interacción estadística.

Una valoración adecuada sobre la existencia de diferencias entre dos subgrupos requiere la consideración de la interacción estadística o test de interacción $[p(i)]$. El análisis estadístico de interacción establece la diferencia de beneficio entre subgrupos mediante el cálculo de $p(i)$, que contribuye a valorar en qué medida la diferencia encontrada puede ser compatible con el azar. Es decir, $p(i)$ mide si las diferencias entre dos o más subgrupos son compatibles con el azar, o se debe al efecto de un fármaco o intervención sanitaria (10,12). $p(i)$ debe ser estimado para aceptar una diferencia de efecto entre subgrupos. Las diferencias estadísticamente significativas entre subgrupos son también denominadas como heterogeneidad. En este aspecto del análisis no se atiende a resultados intragrupal, es decir, a comprobar si un determinado subgrupo presenta o no un beneficio

“estadísticamente significativo” con respecto al valor neutro. En cambio, se atiende a las comparaciones intergrupales, al verificar si existe indicio de diferencia en el resultado de los subgrupos, comparados entre sí (4,8,10,12). Considerando que al dividir el tamaño muestral global del estudio en subgrupos se limita la potencia estadística, se acepta $p(i) < 0.1$ como un valor indicativo de diferencia estadísticamente significativa [o $p(i) \leq 0.2$ si la regresión es multivariante y pre-especificada]. Una mayor flexibilidad sobre el valor de $p(i)$ en este último caso se debe a que la regresión multivariante es un método estadístico que nos permite establecer una relación entre un conjunto de factores y el resultado final, lo cual podría “atenuar” las limitaciones del análisis por subgrupos. Por otra parte, la pre-especificación será descrita a continuación.

- Pre-especificación de los subgrupos.

Asimismo, la aparición de diferencias entre subgrupos en estos estudios debe ser cuidadosamente analizadas para no aceptar los mencionados “artefactos”. Un abundante número de determinaciones podría generar la aparición de diferencias que realmente no son reales por una multiplicidad excesiva. Por ejemplo, si se desarrollaran análisis por subgrupos de acuerdo a 20 características diferentes de la población (género, edad, mutaciones genéticas, etc.) habría más del 50% de probabilidad de encontrar un falso positivo (13). Por ello, es importante que el análisis se centre en unos pocos factores pre-especificados en el protocolo del estudio (14), ya que la probabilidad de encontrar diferencias que realmente son inexistentes (error α) aumenta considerablemente con el número de análisis efectuados. Este contexto podría ser aprovechado por instituciones u organizaciones interesadas para realizar conclusiones *ad hoc* o *post hoc*, favoreciendo la comercialización y uso de fármacos ineficaces o inseguros. Estas conclusiones *ad hoc* o *post hoc* suelen ser interpretaciones de análisis que no estaban protocolizadas en el estudio, que se llevan a cabo por la aparición de resultados favorables inesperados que pueden deberse al azar, y no a una relación de causalidad. Teniendo en cuenta esto, parece razonable pensar en la importancia de la pre-especificación de los factores evaluados en el análisis por subgrupos, que suelen ser aquellos que pueden tener influencia en los resultados obtenidos por la intervención sanitaria. Un ejemplo podría ser que, debido al

carácter mutagénico del tabaco y su impacto negativo en la salud, podría pre-especificarse el análisis del efecto de inmunoterapia en los pacientes fumadores y no fumadores con cáncer de pulmón (15). El tabaco podría generar alteraciones genéticas en los pacientes fumadores que pueden pre-disponer a esta subpoblación para una mayor –o menor– eficacia de la inmunoterapia en comparación con los pacientes no fumadores.

- Plausibilidad biológica.

La plausibilidad biológica es un criterio adicional que nos permite disminuir el riesgo de asumir diferencias inexistentes (error α) al aplicar los resultados de un análisis por subgrupos (1). El factor evaluado que subdivide a la población global del estudio, y que está implicado en la posible heterogeneidad, debe poder ser considerado como causa del efecto observado mediante una hipótesis plausible que permita explicar las diferencias estadísticas encontradas, y a ser posible, enunciada previamente en estudios previos. La heterogeneidad (diferencias) entre subgrupos tendrá mayor credibilidad si tiene un fundamento que lo apoye: fisiopatología, racionalidad biológica, mecanismo de acción del fármaco, etc. (4). Un ejemplo puede ser la observación de un resultado de mayor eficacia de pembrolizumab en pacientes con cáncer de pulmón y una elevada expresión del ligando de muerte programada 1 (PD-L1, del inglés *programmed death ligand 1*), ya que este fármaco actúa sobre este biomarcador (16). Por tanto, la hipótesis de un mayor efecto de este anticuerpo en subgrupos con una mayor expresión de PD-L1 es razonable, considerándose los resultados del análisis por subgrupos y se ha usado como hipótesis plausible para realizar ensayos clínicos confirmatorios (17). De esta manera, los análisis de subgrupos con efectos no explicables por el conocimiento biológico existente suelen ser irreales frecuentemente.

- Consistencia.

El último de los criterios mencionados es la consistencia, que permite verificar la presencia de la misma heterogeneidad o diferencia entre subgrupos en estudios similares. Su valoración precisa de una búsqueda de estudios comparables, en los que pueda entrar en juego la misma hipótesis que sostiene la relación causal, otras variables de resultados relacionadas, estudios del mismo fármaco en otra indicación, o bien fármacos con mismo mecanismo de acción. Encontrar consistencia entre los resultados de los subgrupos en

estos estudios supone un gran apoyo para la fiabilidad de las diferencias encontradas, y al contrario cuando se observa inconsistencia. Meta-analizar los resultados por subgrupos de estudios similares puede aportar mayor información (18). Por ello, un mismo comportamiento de los subgrupos en ECA del uso de un fármaco en una patología podría reforzar la ausencia o la presencia de una diferencia estadísticamente significativa descrita de $p(i)$. Este criterio –junto al de interacción estadística– es esencial en la aplicación de los análisis por subgrupos, ya que evalúa los datos estadísticos obtenidos y balancea las limitaciones del análisis por subgrupos (incremento de los errores α y β).

Además de los aspectos detallados, existe controversia en cuanto a si para considerar los resultados de un análisis de subgrupos es necesario que la característica diferencial se use para estratificar la aleatorización. Sobre esto, es importante destacar que, aunque el efecto de la aleatorización se mantiene en los subgrupos, y que el hecho de que el factor especificado para estratificar la aleatorización imposibilita un tamaño muestral de las ramas muy diferente, no se evita el riesgo de obtener ramas desbalanceadas (14,19). No obstante, en la literatura existen afirmaciones contradictorias (20). El desbalanceamiento de las ramas ocurre cuando pacientes con una característica influyente en el resultado aparecen en mayor medida en una rama que en otra. La probabilidad de generar subgrupos con ramas desbalanceadas respecto a factores relacionados con los resultados es mayor cuanto más reducido sea el tamaño muestral de dichos subgrupos, favoreciendo la aparición de interacciones no fiables. El sesgo producido por el desbalance de las ramas puede reducirse con la aplicación de un análisis multivariante que incluya los factores más relevantes asociados al beneficio. Esta sería la mejor forma de realizar un análisis de subgrupos. Lamentablemente, esta metodología resulta no es frecuente, o no llega a publicarse (21).

Si bien los criterios mencionados han sido descritos en la literatura (4), en la actualidad no existe un consenso sobre su uso a la hora de realizar afirmaciones sobre los resultados de los análisis por subgrupos. Y esto a pesar de que la propia Agencia Europea del Medicamento ha recalcado la relevancia del análisis por subgrupos (22). Los graves errores metodológicos cometidos al interpretar erróneamente los ensayos clínicos, contribuyen a la controversia que envuelve al análisis por subgrupos, dificultando llegar a decisiones homogéneas y bien justificadas sobre selección de medicamentos (4,23,24). La falta de

consenso sobre cómo abordar el análisis por subgrupos favorece la toma de decisiones clínicas basada en criterios superfluos. La interpretación de ausencia de beneficio de una intervención en un subgrupo de pacientes por falta de "significación estadística" del resultado en dicho subgrupo es el error más grave (9). La causa de este y otras interpretaciones inadecuadas consiste en ignorar el imprescindible análisis de interacción.

Existe bibliografía que respalda que no se ha mejorado la interpretación del análisis por subgrupos, a pesar de la importancia que las agencias reguladoras dan a esta entidad (8, 22). De hecho, se han realizado una interesante publicación sobre la falta de datos e incorrecta valoración del análisis por subgrupos (25). En ese estudio se afirma que el 71.8% de las publicaciones –incluso en las revistas científicas de mayor prestigio– ni siquiera refieren el resultado del test de interacción (8, 25). Teniendo en cuenta el enorme impacto clínico, fármaco-económico y social, estos problemas sobre la valoración del análisis por subgrupos son especialmente preocupantes en los ECA específicos de oncología y hematología. La falta de pre-especificación de los análisis por subgrupos, un exceso de factores analizados y un uso inadecuado del $p(i)$ han sido reseñados por una revisión como los defectos metodológicos más importantes (6). De hecho, sólo el 34-18% de los ECA de oncología y hematología evaluados reportaron una valoración del efecto de las terapias en los subgrupos mediante $p(i)$. Todo ello no sólo tiene una repercusión clínica individual en los pacientes, sino que el uso inadecuado de nuevas terapias con enorme impacto económico –como los tratamientos onco-hematológicos– impide el uso racional de los recursos del sistema sanitario.

Tanto es así, que la Organización Mundial de la Salud (OMS) emitió un informe en 2018 alertando del excesivo coste de los tratamientos onco-hematológicos con respecto al resto de ámbitos sanitarios (26). La **Figura 2** puede ser consultada para entender el enorme impacto económico que suponen estas terapias. Este desbalanceamiento del uso de los recursos no hace más que favorecer la inequidad a la hora de prestar asistencia sanitaria a los pacientes de diferentes patologías.

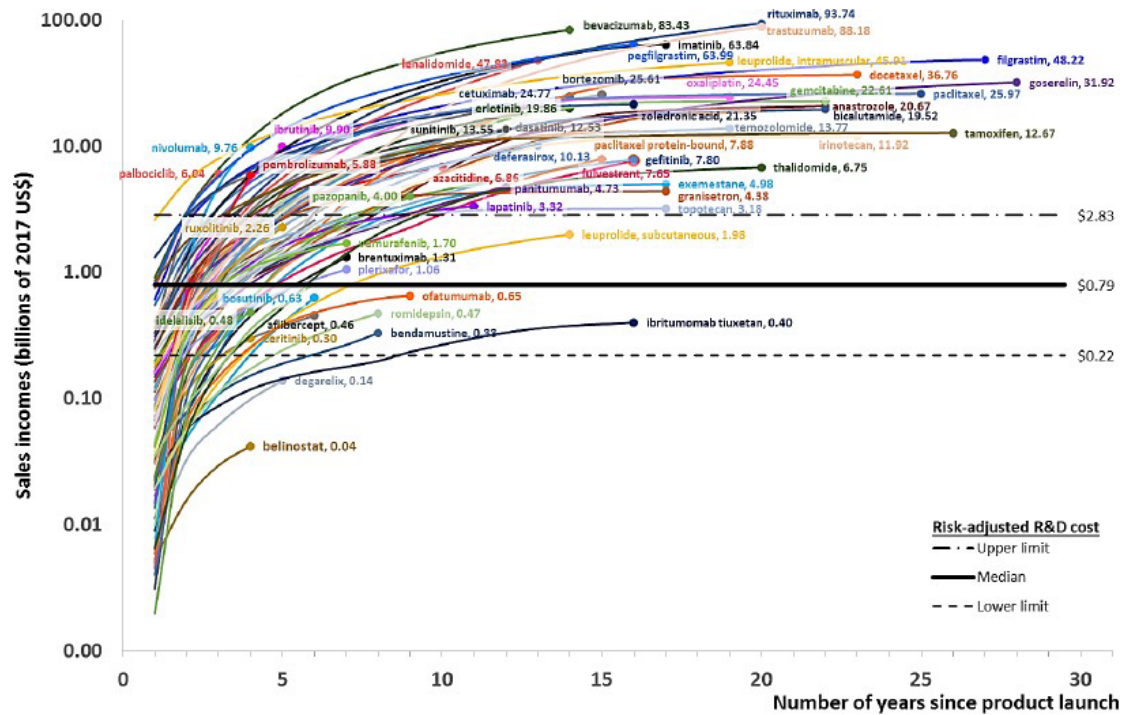


Figura 2. Ingresos acumulados por ventas de medicamentos contra el cáncer en dólares estadounidenses de 2017, por molécula. Tomada de World Health Organization. Technical report: pricing of cancer medicines and its impacts (2018).

El creciente desarrollo de nuevas moléculas conduce a una mayor competencia de los fármacos onco-hematológicos. Pacientes con el mismo diagnóstico y similar perfil clínico pueden ser candidatos a varias moléculas (27–29). En estas situaciones, los resultados obtenidos en los análisis por subgrupos adquieren cada mayor relevancia y las empresas farmacéuticas intentan sacar ventaja frente a sus competidores. Por ello, las afirmaciones sobre los análisis por subgrupos formuladas en los actuales y futuros ECA pueden presentar un impacto considerable en la eficacia, seguridad, conveniencia de los tratamientos onco-hematológicos, eficiencia de los sistemas públicos de salud y posicionamiento terapéutico de las diferentes terapias en un determinado contexto clínico.

No obstante, este escenario no sólo se plantea en las terapias antineoplásicas. La necesidad de una interpretación adecuada del análisis por subgrupos también es crucial en tratamientos cuya población diana engloba a un número muy elevado de pacientes (30). Por otra parte, la irrupción del síndrome de infección respiratoria severa aguda del coronavirus 2 (SARS-CoV-2) y la consecuente emergencia sanitaria ocasionada por la

pandemia por enfermedad por coronavirus 2019 (COVID-19) ha generado multitud de afirmaciones positivas sobre fármacos con un impacto económico importante (31), sin estar fundamentadas sobre una base científica sólida. Algunos medicamentos que pueden servir como ejemplos son remdesivir y la combinación lopinavir/ritonavir. Esto ha generado un consumo de los recursos que ponen en peligro la sostenibilidad del sistema sanitario público.

2. JUSTIFICACIÓN Y OBJETIVOS

2. JUSTIFICACIÓN Y OBJETIVOS

Teniendo en cuenta todos estos antecedentes, podemos afirmar que existe una necesidad de consenso para la aplicación de criterios homogéneos que nos permitan evaluar adecuadamente el análisis por subgrupos. La implantación de una metodología específica para su interpretación podría disminuir las discrepancias y errores de aplicación, y mejorar la lectura crítica por parte de los profesionales sanitarios (1). De esta manera, se podría proporcionar un punto de partida para el intercambio de opiniones en la selección de pacientes candidatos a determinados tratamientos.

El desarrollo de herramientas metodológicas para facilitar una adecuada evaluación, selección y posicionamiento de medicamentos es una labor esencial. De hecho, la aparición de nuevas herramientas que permitan una evaluación ordenada y sistemática del análisis por subgrupos supondría una ventaja a la hora de abordar los datos proporcionados por los ECA. La revisión de los análisis por subgrupos según los criterios objetivos posibilitaría reducir la incertidumbre en la toma de decisiones clínicas (4). La consideración de $p(i)$, el número de subgrupos evaluados, la existencia de una hipótesis descrita que justifique una diferencia entre subgrupos o la consistencia de los resultados entre estudios similares son algunos de estos criterios.

Su aplicación permitiría reducir el incremento de errores α y β inherente al análisis por subgrupos, evitando considerar falsas diferencias aparentes entre subpoblaciones o ser demasiado restrictivo al tener en cuenta criterios clásicos estadísticos. Esto no sólo facilitaría a los expertos el mencionado intercambio de opiniones teniendo en cuenta criterios estandarizados, sino que ayudaría a profesionales no familiarizados con el análisis por subgrupos a seguir unas pautas de valoración. De esta manera, analizar metodológicamente las ventajas e inconvenientes de fármacos en pacientes con características diferenciadas reflejadas en los análisis de subgrupos podría favorecer el posicionamiento terapéutico de las terapias, evitando los importantes problemas metodológicos que sesgan la interpretación de los datos y conducen a decisiones erróneas (2,23,24).

A lo largo de este trabajo se pretende elaborar una herramienta estructurada que facilite la interpretación de los análisis por subgrupos y la evaluación de su aplicabilidad, aportando criterios claros sobre los que debe pivotar el análisis, para favorecer el proceso de diálogo y la toma de decisiones de posicionamiento terapéutico. Asimismo, se aplicará la metodología desarrollada en contextos clínicos específicos, y se evaluará el impacto clínico y fármaco-económico del análisis por subgrupos en estos contextos clínicos.

Por otra parte, la realización de análisis farmacoeconómicos sobre los datos de subgrupos es esencial para demostrar su enorme impacto. Este hecho resulta claramente palpable en fármacos con un coste considerable que, además, pueden ser usados en un número de pacientes elevado. Este es el caso de mepolizumab en el tratamiento del asma eosinofílica refractaria grave

Los **objetivos específicos** que se pretenden alcanzar en la presente Tesis Doctoral son los siguientes:

1. Diseño y validación de un cuestionario con una metodología sistemática para el abordaje del análisis por subgrupos.
2. Aplicación de la herramienta desarrollada en contextos clínicos de elevado impacto clínico, fármaco-económico y social. Este apartado contendrá tanto revisiones sistemáticas de análisis por subgrupos en determinados contextos clínicos, como publicaciones concretas.
3. Desarrollo de un estudio fármaco-económico sobre mepolizumab en el tratamiento del asma eosinofílica refractaria grave, considerando los datos de análisis por subgrupos.

3. RESULTADOS

3.1. Desarrollo de una herramienta para la interpretación y aplicación del análisis por subgrupos.

Gil-Sierra MD, Fénix-Caballero S, Abdel Kader-Martin L, Fraga-Fuentes MD, Sánchez-Hidalgo M, Alarcón de la Lastra-Romero C, et al. Checklist for clinical applicability of subgroup analysis. *J Clin Pharm Ther.* 2020;45(3):530-8. Doi:10.1111/jcpt.13102

3.1. DESARROLLO DE UNA HERRAMIENTA PARA LA INTERPRETACIÓN Y APLICACIÓN DEL ANÁLISIS POR SUBGRUPOS.

3.1.1. Introducción

El desarrollo de herramientas que permitan un abordaje objetivo, sistemático y ordenado del análisis por subgrupos es esencial para una evaluación homogénea de esta entidad. De esta manera se evitaría la consideración de datos superfluos que carecen de relevancia clínica. El objetivo de este apartado es la descripción del diseño y validación de un cuestionario que considera los criterios de interacción estadística, pre-especificación, plausibilidad biológica y consistencia de resultados en el análisis por subgrupos.

3.1.2. Metodología

En el presente estudio, se tuvo en cuenta bibliografía previa relativa a la metodología existente para la interpretación y aplicación del análisis por subgrupos, con una posterior síntesis de la información recopilada (2,4,5,9,10,12,19,20,24). Teniendo en cuenta esta información, se establecieron una serie de cuestiones preliminares de descarte para estimar la idoneidad y validez del estudio (**Figura 3**). Estas cuestiones preliminares fueron las siguientes: 1) consideración del nivel de evidencia y aplicabilidad del estudio valorado, así como la relevancia de la variable evaluada en el análisis por subgrupos, 2) selección – atendiendo a $p(i)$ o heterogeneidad– de subgrupos en los que se observan posibles diferencias de efecto, y 3) valoración de la secuencia temporal entre el factor que establece los subgrupos y el efecto (32).

Posteriormente, se elaboró el cuestionario propiamente dicho, el cual estuvo compuesto de tres apartados diferenciados que recogieron ordenadamente los diferentes factores involucrados en la fiabilidad de la posible diferencia observada entre subgrupos: I) asociación estadística, que englobó los criterios de interacción estadística [$p(i)$] y pre-especificación de subgrupos evaluados; II) plausibilidad biológica y III) consistencia de resultados. Cada apartado se dividió a su vez en una serie de cuestiones y sus correspondientes respuestas. Cada respuesta se relacionó con un nivel de fiabilidad, que

podía adquirir las valoraciones de probable, posible, dudosa y nula. Para dar respuesta a cada apartado y facilitar la aplicación del cuestionario, se incluyeron aclaraciones y definiciones de determinados conceptos.

La asociación estadística consideró el valor de $p(i)$, estableciendo los límites en $p(i) < 0.05$ (12), $0.05 \geq p(i) < 0.1$ (2) o ninguno de los anteriores, y la pre-especificación de los subgrupos evaluados. Como ya se ha mencionado anteriormente, la división del tamaño muestral total de un ECA en subgrupos disminuye la potencia estadística. Por ello, se aceptó un valor $p(i) < 0.1$ como indicativo de diferencia estadísticamente significativa (o $p(i) \leq 0.2$ se ha desarrollado una regresión multivariante y pre-especificada). Para calcular el valor de $p(i)$ entre dos subgrupos cuando los estudios no aportan su estimación, se usó una calculadora basada en la escala de logaritmo neperiano (\ln) de los valores de riesgo relativo, odds ratio, hazard ratio e ICs proporcionados por los ensayos clínicos (33). Asimismo, el tamaño de la población de subgrupos (superior o inferior a 50), el número de factores pre-especificados evaluados (superior o inferior a 10) y el resultado global del estudio (positivo o negativo) constituyeron cuestiones adicionales integradas en el criterio de asociación estadística. De acuerdo a la literatura previa, la fiabilidad fue menor para los estudios con resultado global negativo en los que se considerase un posible efecto en un subgrupo (11).

El apartado sobre plausibilidad biológica consideró la existencia de una hipótesis que justificara una relación de causa-efecto para la posible diferencia encontrada entre los subgrupos, diferenciando entre hipótesis descritas previamente en literatura con base científica razonable, aquellas hipótesis que carecen de una base previa, las dudosas y los hallazgos que carecen de hipótesis plausible. En el apartado de consistencia de resultados, se valoró disponer de interacción estadística en estudios similares o meta-análisis, y en su defecto, la presencia de resultados compatibles. Este último escenario es aquel en el que los resultados no muestran interacción estadística o heterogeneidad en todos los ensayos, pero sí en el mismo sentido. Se consideró también el escenario en el que se carece de estudios comparables y el contexto con inconsistencia de resultados.

Se asignó una puntuación determinada a cada valoración de nivel de fiabilidad: probable (+3 puntos), posible (+2), dudosa (0) y nula (-3). Se evaluó el análisis por subgrupos

mediante el sumatorio global de las puntuaciones siempre que la asociación estadística o la consistencia de resultados no fueran nulas. Si algunos de estos dos apartados fueran nulos, se descartó directamente la aplicación del análisis por subgrupos. El sumatorio global se relacionó con una recomendación orientativa para la aplicabilidad del resultado de un subgrupo en la toma de decisiones clínicas: probable (9-7 puntos) correspondió con la recomendación de aplicabilidad del análisis por subgrupos mientras no se disponga de ECA confirmatorio; posible (5-6 puntos) con prudente aplicabilidad en casos de elevada toxicidad, dificultad de uso o precio de las alternativas terapéuticas; dudosa (3-4 puntos) con rechazo de la aplicabilidad salvo situaciones excepcionales; y nula (<3 puntos) que sustentó el rechazo absoluto de la posible diferencia entre subgrupos para el posicionamiento terapéutico.

La herramienta diseñada fue enviada a los miembros del Grupo de Evaluación de Novedades, Estandarización e Investigación en Selección de Medicamentos (GÉNESIS) de la Sociedad Española de Farmacia Hospitalaria. Este grupo de trabajo está compuesto por expertos en evaluación, selección y posicionamiento terapéutico de medicamentos. Estos profesionales estuvieron de acuerdo con el contenido del manuscrito enviado y realizaron aportaciones. Aquellas contribuciones razonables fueron estudiadas e incorporadas.

Tras el diseño del cuestionario, se llevó a cabo su validación. Para ello, se evaluó la concordancia inter-investigador mediante la evaluación por profesionales sanitarios de tres análisis por subgrupos propuestos sobre ramucirumab (34,35) (caso 1), nivolumab (36) (caso 2) y mepolizumab (30,37) (caso 3). Cada uno de los casos se analizó usando individualmente la herramienta antes (análisis *pre*) y después (análisis *post*) de una discusión en común sobre la aplicabilidad de los datos de los subgrupos. Esto se hizo para comprobar la influencia de una puesta en común de las evaluaciones para mejorar la concordancia de los resultados, tal como se realiza en otras herramientas (38).

Asimismo, se registró el ámbito de los profesionales sanitarios participantes, su experiencia previa en evaluación y posicionamiento terapéutico de fármacos. Finalmente, se realizaron dos preguntas a cada investigador sobre su recomendación acerca de la aplicación del análisis por subgrupos en el caso estudiado (I) y la utilidad del cuestionario para evaluar la

aplicabilidad de las posibles diferencias por subgrupos (II): I) ¿Recomendaría usted aplicar en la práctica clínica la posible diferencia encontrada en los resultados por subgrupos? Respuesta: Sí/No, II) ¿Le ha servido esta herramienta para evaluar la aplicabilidad clínica de este análisis por subgrupos? Respuesta: 1-5. La mayor utilidad se corresponde con una puntuación de 5 (mucho) y la de menor utilidad con una puntuación de 1 (nada).

La validación de la herramienta se realizó comprobando la concordancia inter-investigador mediante el valor kappa (κ) para aplicabilidad global, así como en los apartados de asociación estadística, plausibilidad biológica y consistencia. Se emplearon los indicadores de concordancia simple observada (A_o), definido como la concordancia en la respuesta más frecuente, y de concordancia esperada (A_e), definido como la concordancia esperada según el número de respuestas posibles en cada cuestión. K fue calculada usando la siguiente fórmula: $\kappa = (A_o - A_e) / (1 - A_e)$. La fuerza de la concordancia se interpretó usando los siguientes valores de κ (39): <0.2 se relacionó con una pobre concordancia; 0.21-0.4 con débil; 0.41-0.6 con moderada; 0.61-0.8 con buena y 0.81-1 con muy buena concordancia. El test chi-cuadrado se utilizó para comparar los coeficientes kappa de “asociación estadística”, “plausibilidad biológica”, “consistencia” y “aplicabilidad” debido a que estas variables son variables cualitativas. Asimismo, las recomendaciones de aplicación de los resultados de los subgrupos (respuestas posibles: sí/no) también se analizaron mediante el test chi-cuadrado. No obstante, la utilidad del cuestionario fue usada en el análisis mediante la prueba no paramétrica de la U de Mann-Whitney debido a que esta variable cuantitativa presentó distribución no normal.

Se presentó la posibilidad de realizar recomendaciones para la mejora del cuestionario a los profesionales que intervinieron en la fase de validación (al igual que se hizo en la etapa de diseño con los componentes del grupo GÉNESIS), aceptándose aquellas propuestas que fueron razonables.

3.1.3. Resultados

La **Figura 3** es la versión final de la herramienta, la cual consta de dos partes diferenciadas. La primera parte (A) muestra las cuestiones preliminares detalladas anteriormente. Esta primera parte aborda que la valoración de la aplicabilidad de la posible diferencia encontrada en los subgrupos requiere que el estudio presente la mayor evidencia disponible, que se haya evaluado una variable de clara relevancia clínica o seguridad, y que los pacientes presenten el factor diferenciador de los subgrupos antes de recibir la intervención. La segunda parte (B) está constituida por el propio cuestionario, que incluye siete preguntas sobre los tres apartados mencionados con anterioridad: cinco hacen referencia a la asociación estadística, una a la plausibilidad biológica y otra a la consistencia. Las cuestiones sobre la asociación estadística presentan once posibles respuestas, con sus correspondientes grados de fiabilidad del análisis. Las cuestiones sobre plausibilidad biológica y consistencia tienen cuatro posibles respuestas cada una, relacionadas también con sus diferentes grados de fiabilidad.

Se dispuso de 26 participantes de diferentes centros en el proceso de validación de la herramienta: 15 (57.7%) fueron farmacéuticos hospitalarios, 5 (19.2%) fueron farmacéuticos de atención primaria, 5 (19,2%) se dedicaban al ámbito de administración sanitaria y 1 (3.8%) fue un profesional de la industria farmacéutica. En cuanto a la experiencia previa en evaluación y posicionamiento de fármacos, las características de los usuarios de la herramienta fueron: 15 (57.7%) profesionales evaluaban fármacos habitualmente, 4 (15.4%) de forma ocasional y 7 (26.9%) carecían de experiencia evaluadora.

Figura 3. Herramienta para valorar la aplicabilidad de un análisis de subgrupos en la toma de decisiones de la práctica clínica.

Este cuestionario está enfocado para la valoración del resultado de una intervención frente a un control en subgrupos que aparentemente muestran efectos diferentes. Aporta puntuaciones meramente orientativas, útiles para ayudar en el proceso de análisis y decisión. La aplicación del cuestionario no sustituye al buen criterio de los evaluadores, y puede necesitar adaptaciones o correcciones al aplicarse a los diversos escenarios y circunstancias que pueden presentarse en la práctica clínica.

A) Cuestiones preliminares

Antes que atender a cualquier otro criterio, hay que seguir los tres puntos siguientes:

1. Aplicar unas cuestiones de descarte para valorar si el análisis reúne las condiciones básicas para tener impacto en la toma de decisiones de la práctica clínica.

El estudio sobre el que se está realizando el análisis:

- a. Presenta el mayor nivel de evidencia y aplicabilidad disponible para justificar la práctica asistencial: continuar.
- b. Presenta un nivel de evidencia o aplicabilidad insuficiente para la toma de decisiones terapéuticas: el análisis de subgrupos no debe usarse para el posicionamiento terapéutico, exista o no interacción.

La variable considerada es:

- a. Una variable de clara relevancia clínica o una variable subclínica o intermedia, la cual ha sido medida como variable principal en el estudio. Si esta variable resulta aceptable para la toma de decisiones terapéuticas en el escenario clínico de referencia: continuar.
- b. Una variable relevante de seguridad: continuar
- c. En el resto de casos: el análisis de subgrupos no debe usarse para el posicionamiento terapéutico, se muestre o no interacción.

2. Seleccionar los factores que determinan subgrupos con diferencias de efecto entre sí. Esto se contempla por la diferencia o desplazamiento de sus valores e intervalos de confianza, mediante $p(i)$. Una $p(i)$ significativa ($<0,1$ o menor de un valor razonable pre-especificado por los autores del estudio) indica que ese factor debe ser valorado. En ausencia de una determinación de $p(i)$, intervalos que no se solapan o lo hacen escasamente (el valor extremo de uno no llega al valor central del otro) indican que podría existir heterogeneidad y podríamos aplicar el cuestionario, calculando la $p(i)$ si es posible. El cuestionario se aplica para cada factor seleccionado. Como ya hemos dicho, un factor seleccionado es aquél que discrimina subgrupos que parece que muestran resultados diferentes entre sí.

- Una excepción es cuando el factor suponga a priori diferencias muy importantes entre los subgrupos de población, de forma que se justificarían ensayos diferentes. En tal caso, puede ser razonable exigir una confirmación estadísticamente significativa del resultado global en cada subgrupo por separado. Por tanto, no se aplicaría el cuestionario.

3. Considerar la secuencia temporal:
 - a. Los pacientes presentaban el factor que establece los subgrupos antes de recibir la intervención: continuar.
 - b. Los pacientes presentaban el factor que establece los subgrupos después de empezar la intervención, pudiendo tratarse de un simple marcador de riesgo sin relación con el efecto de la intervención: no continuar.

B) Valoración de la aplicabilidad de la posible diferencia encontrada.

Se valoran las cuestiones reflejadas en la siguiente tabla.

(La figura continúa en la siguiente página)

Criterio	Cuestión	Respuestas posibles	Fiabilidad del análisis	Orientaciones
I-Asociación estadística	Cuestiones esenciales			
	1. Interacción: ¿Existe interacción o heterogeneidad, es decir, diferencia estadísticamente significativa entre el resultado de los distintos subgrupos?	Sí, con $p(i) < 0,05$.	Probable	Se considera significativo un valor de $p(i) < 0,05$. Como la división en subgrupos disminuye la potencia estadística, se puede considerar indicativo $p(i) < 0,1$, o incluso $p(i) < 0,2$ en regresión multivariante, si los autores lo especifican. En muchos estudios no se aporta $p(i)$. No obstante, es posible calcularla. Una $p(i)$ reducida o un análisis de interacción multivariante disminuyen la incertidumbre sobre la asociación, especialmente cuando se han analizado muchos factores distintos. <i>(*) Se admitiría una excepción a la exigencia de interacción para considerar los resultados de un subgrupo: cuando por un defecto de diseño, el estudio mezcla poblaciones muy heterogéneas con respecto al efecto esperable, que justificarían estudios distintos. En este caso, podría requerirse un resultado confirmatorio en cada subgrupo.</i>
		Sí, con $p(i) \geq 0,05$ pero $< 0,1$ (o $p(i) < 0,2$ en regresión multivariante si así se pre-especifica)	Dudosa	
		Ninguna de las anteriores	Nula(*)	
	2. Pre-especificación: ¿Estaba pre-especificado en la metodología estudiar ese factor?	Sí	Probable	Es fácil encontrar un grupo que presente diferencias <i>post hoc</i> . Por eso, es esencial que el subgrupo estudiado estuviera pre-especificado para su análisis. En su defecto, se puede considerar que el subgrupo fue pre-especificado si fue elegido para estratificar la aleatorización.
		No, o no se conoce	Nula	
	Cuestiones adicionales			
	3. Población (N) de los subgrupos	N >50/rama (>100 en total)	Probable	Una N demasiado reducida facilita la descompensación de las características basales en las ramas estudiadas.
		N <50 en alguna rama	Dudosa	
	4. Número de factores estudiados: contar los factores pre-especificados /de estratificación (*) analizados por subgrupos.	1-9 factores pre-especificados analizados	Probable	Cuantos más factores se analicen, mayor es la probabilidad de encontrar heterogeneidad azarosa o por ramas descompensadas en factores pronósticos, sobre todo si la $p(i)$ no es muy reducida. <i>(*) Si el factor es pre-especificado, contar n° de factores pre-especificados. Si es usado en la estratificación, pero no es pre-especificado (caso raro): contar preespecificados + estratificados. Si el factor no es pre-especificado, asumir siempre >10, independientemente de los que aparezcan.</i>
		10 o más factores	Dudosa	
	5. Resultado global del estudio	Positivo, pero podría haber diferencias entre subgrupos.	Probable	Cuando no se ha podido demostrar el beneficio según el diseño planificado del estudio, el análisis de subgrupos por sí solo tiene una fiabilidad limitada.
		Negativo, pero podría ser positivo en subgrupos	Dudosa	
	Valoración global del criterio "asociación estadística"	Todas respuestas "probables"	Probable	La valoración de "asociación estadística" se integra luego con las valoraciones de "plausibilidad biológica" y "consistencia"
1 "dudosa", 0 "nulas"		Posible		
2 "dudosas", 0 "nulas"		Dudosa		
≥ 3 "dudosas" o alguna "nula"		Nula		

(La figura continúa en la siguiente página)

criterio	Cuestión	Respuestas posibles	Fiabilidad del análisis	Orientaciones
II- Plausibilidad biológica	¿Existe una hipótesis plausible para justificar la relación de causa y efecto?	Sí, basada en literatura previa	Probable	Las diferencias deben poder ser sustentadas por una hipótesis aceptable, acorde al conocimiento actual de la enfermedad. Las bases para esta hipótesis deben estar presentes en la literatura. Que se haya considerado dicho factor en la estratificación muestra que se preveía su posible influencia en el resultado. Cuando un factor cuantitativo divide la población en tres o más subgrupos, resultados compatibles con un gradiente de efecto -también conocido como efecto "dosis-respuesta"- apoyarían la hipótesis, y al contrario.
		Se puede enunciar una hipótesis razonable, pero sin base en literatura previa	Posible	
		Se puede enunciar una hipótesis dudosa, pero sin base en literatura previa	Dudosa	
		Sin hipótesis plausible	Nula	
III- Consistencia	¿Existe consistencia con resultados obtenidos en otros estudios similares?	Sí, estudios similares muestran heterogeneidad, o aparece interacción en meta-análisis y los estudios son homogéneos	Probable	Si existen estudios similares, los resultados en los grupos de pacientes estudiados deben ser compatibles con el hallazgo analizado. Estudios "similares" son aquellos desarrollados en la misma patología, con la misma intervención y control. Si en dos o más estudios se observa heterogeneidad y no hay resultados contradictorios, la consistencia es probable. En ausencia de estudios similares, se pueden analizar otros con fármacos o patologías relacionadas, siempre que se estime que en ellos debería cumplirse la misma hipótesis que apoya la plausibilidad biológica. En tal caso, la consistencia tendrá valoración máxima de "posible". Si no hay otros estudios con los que comparar, se puede comprobar si existe falta de consistencia interna con los resultados de otras variables relevantes relacionadas con la estudiada. Es difícil, por el aumento del error beta debido a la reducción del tamaño muestral, que en todos los estudios aparezca interacción significativa. Puede recurrirse a meta-análisis de resultados por subgrupos y, una vez observada la interacción en el análisis global, comprobar su homogeneidad, que apoyaría la diferencia (valoración de "probable").
		No, pero existen resultados compatibles (sin interacción en todos, pero en el mismo sentido).	Posible	
		No existen otros estudios con los que comparar. No existe, al menos, falta de consistencia interna.	Dudosa	
		No se cumple ninguna de las condiciones anteriores	Nula	

(La figura continúa en la siguiente página)

Valoración global de la aplicabilidad

- 1) Si la asociación estadística o la consistencia reciben valoración de “nulas”, la aplicabilidad sería nula.
- 2) En el resto de casos, la siguiente puntuación puede resultar orientativa. En todo caso, es opcional, y no sustituye al buen juicio de los evaluadores. Se puntúa cada una de las 3 valoraciones en “asociación estadística”, “plausibilidad biológica” y “consistencia”. Los resultados podrían variar -al alza o a la baja- teniendo en cuenta las excepciones comentadas u otras apreciaciones razonables de los evaluadores.

Criterio	Valoración probable, posible, dudosa o nula	Puntuación Probable: 3 puntos. Posible: 2 puntos. Dudosa: 0 puntos. Nula: -3 puntos.
Asociación estadística		
Plausibilidad biológica		
Consistencia		
Aplicabilidad práctica (sumatorio): probable (7-9), posible (5-6), dudosa (3-4), nula (<3).		

Recomendaciones para considerar los resultados de análisis por subgrupos en la toma de decisiones clínicas.

Aplicabilidad	Recomendación
Probable	Los resultados del análisis de subgrupos deberían ser tenidos en cuenta en la práctica clínica, hasta que no se disponga de un ensayo clínico diseñado específicamente para confirmar o descartar la diferencia de efecto.
Posible	Los resultados del análisis por subgrupos podrían ser aplicables a la práctica clínica con prudencia, si la situación lo aconseja. Deberían ser tenidos en cuenta si la intervención resulta más tóxica, difícil de utilizar o muy cara con respecto a sus alternativas.
Dudosa	El análisis de subgrupos no debe ser tenido en consideración para la práctica clínica. *Excepcionalmente, podría sopesarse su aplicación en el caso de una intervención mucho más tóxica, muy difícil de usar o de coste inasumible. De esta manera, se aconsejaría restringir el uso a situaciones de beneficio indudable.
Nula	El análisis de subgrupos no muestra diferencias que puedan ser consideradas de forma fiable en la práctica clínica. Los resultados por subgrupos no deben usarse para valorar la eficacia, ya que el resultado atribuible a cada subgrupo no es otro que el global del estudio.

Nota. Consideración adicional: diferencias en riesgo absoluto

Los análisis habituales de subgrupos se limitan a estudiar diferencias en el efecto, generalmente en términos relativos (riesgo relativo, odds ratio, hazard ratio, razón de tasas, etc.). Sin embargo, existiendo o no variaciones relativas en el efecto entre los diferentes subgrupos, pueden existir, además, diferencias importantes en la magnitud del mismo, en subgrupos con diferente riesgo basal (40).

Estas diferencias en la magnitud absoluta del efecto en los diferentes subgrupos pueden tener notables repercusiones y mostrar diferencias en la relevancia clínica del beneficio en las distintas subpoblaciones. En consecuencia, también pueden hacer variar las relaciones beneficio/riesgo y beneficio/coste. Esta circunstancia no ha sido valorada en la presente herramienta, pero puede tener implicaciones relevantes para las decisiones de posicionamiento terapéutico de una intervención, y requeriría otro estudio complementario.

Por otro lado, la **Tabla 1** muestra los datos obtenidos durante la validación de la herramienta. En el análisis *pre*, la concordancia inter-investigador de la aplicabilidad fue $\kappa = 0.79$ (buena) en el caso 1, $\kappa = 1.00$ (muy buena) en el caso 2 y $\kappa = 0.83$ (muy buena) en el caso 3. En el análisis *post*, la concordancia inter-investigador de la aplicabilidad fue $\kappa > 0.90$ (muy buena) en los tres casos. En cuanto a los valores κ de “asociación estadística”, “plausibilidad biológica” y “consistencia”, las menores concordancias inter-investigador del análisis *pre* se obtuvieron en “plausibilidad biológica” del caso 1 ($\kappa = 0.44$; moderada) y en consistencia del caso 2 ($\kappa = 0.36$; débil). Ambos resultados mejoraron en el análisis *post* ($\kappa = 0.83$ para “plausibilidad biológica” del caso 1 y $\kappa = 0.94$ para “consistencia” del caso 2). Las puntuaciones de utilidad de la herramienta fueron 4.77/5 en el caso 1, 4.88/5 en el caso 2 y 4.75/5 en el caso 3. En el análisis *pre*, la concordancia inter-investigador de la recomendación global de aplicabilidad de los resultados por subgrupos en la práctica fue 92.3% (concordancia en no aplicar el resultado por subgrupos) en el caso 1, 96% (no aplicar) en el caso 2 y 100% (favorable a aplicación) en el caso 3. En el análisis *post*, la concordancia inter-investigador de la recomendación de aplicabilidad de los resultados por subgrupos fue 100% (no aplicación) en el caso 1, 94.45% (no aplicación) en el caso 2 y 100% (favorable a aplicación) en el caso 3.

Se realizaron tres modificaciones menores en el cuestionario tras las sugerencias de mejora por parte de los componentes del grupo GENESIS y de los participantes en la fase de validación. Dos correspondieron a modificaciones en las orientaciones del apartado de “consistencia” del cuestionario, para favorecer la comprensión del concepto estudios similares, reducir la subjetividad de las respuestas y orientar a los usuarios en situaciones sin estudios comparables o resultados compatibles. La modificación restante se hizo en las orientaciones del apartado de “asociación estadística”, más concretamente en la cuestión de factores estudiados. Esto se llevó a cabo para facilitar la diferenciación entre factores pre-especificados y factores estratificados.

Tabla 1. Resultados de concordancia en la validación del cuestionario.

Caso 1: ramucirumab					
	Análisis <i>pre</i> (n = 26)		Análisis <i>post</i> (n = 24)		P*
	Ao (%)	κ	Ao (%)	κ	
Asociación estadística	88.46	0.85	91.67	0.89	0.70
Plausibilidad biológica	57.69	0.44	87.5	0.83	0.01
Consistencia	76.92	0.69	95.83	0.94	0.04
Aplicabilidad	84.62	0.79	95.83	0.94	0.22
Mediana	80.77	0.74	93.75	0.92	0.16
Recomendación aplicabilidad	No	Si	No	Si	0.49
	24 (92.3%)	2 (7.67%)	24 (100%)	0	
Utilidad cuestionario†	4.77		4.62		0.44
Caso 2: nivolumab					
	Análisis <i>pre</i> (n = 25)		Análisis <i>post</i> (n = 22)		P*
	Ao (%)	κ	Ao (%)	κ	
Asociación estadística	96	0.95	100	1	0.37
Plausibilidad biológica	96	0.95	95.45	0.94	0.92
Consistencia	52	0.36	95.45	0.94	<0.001
Aplicabilidad	100	1	100	1	-
Mediana	96	0.95	97.72	0.94	0.80
Recomendación aplicabilidad	No	Si	No	Si	>0.1
	24 (96%)	1 (4%)	21 (94.45%)	1 (4.55%)	
Utilidad cuestionario†	4.88		4.72		0.3
Caso 3: mepolizumab					
	Análisis <i>pre</i> (n = 24)		Análisis <i>post</i> (n = 21)		P*
	Ao (%)	κ	Ao (%)	κ	
Asociación estadística	91.67	0.89	90.91	0.88	0.92
Plausibilidad biológica	87.5	0.83	90.91	0.88	0.71
Consistencia	83.33	0.78	100	1	0.03
Aplicabilidad	87.5	0.83	100	1	0.19
Mediana	87.5	0.83	95.45	0.94	0.62
Recomendación aplicabilidad	No	Si	No	Si	-
	0	24 (100%)	0	21 (100%)	
Utilidad cuestionario†	4.75		4.86		0.38

Análisis *pre*: análisis usando el *checklist* de forma individual antes de una valoración común. Análisis *post*: análisis usando el *checklist* de forma individual después de una valoración común. Ao: concordancia en la respuesta más frecuente. κ: valor kappa. †La utilidad del cuestionario fue evaluada mediante el método U-Mann Whitney (test no paramétrico). *El valor de P se refiere a la comparación del análisis *pre* vs análisis *post* de los valores κ obtenidos mediante el test chi-cuadrado. Se analizó la recomendación de aplicabilidad (sí/no) usando el test chi-cuadrado.

3.1.4. Discusión

Se ha llevado a cabo el desarrollo de un instrumento que facilita la interpretación y aplicación del análisis por subgrupos. De esta manera, se han podido establecer unos criterios homogéneos para que los profesionales sanitarios puedan entablar discusiones sobre la evaluación y posicionamiento de nuevos fármacos teniendo en cuenta el análisis por subgrupos. Se podría desechar aquella información que carece de interés y mejorar la interpretación de la evidencia científica por parte de los profesionales sanitarios en general, y de facultativos clínicos en particular. El cuestionario recibió unas excelentes valoraciones de utilidad para los usuarios y mostró una elevada concordancia interindividual de los resultados, tanto para evaluar la aplicabilidad del análisis por subgrupos como para realizar recomendaciones para la práctica clínica. Los casos puntuales que mostraron una concordancia débil/moderada mejoraron tras la discusión en grupo, generándose un escenario similar a las comisiones farmacoterapéuticas de los hospitales o grupos de expertos de las sociedades científicas.

Además de la consideración de bibliografía previa, la herramienta fue desarrollada teniendo en cuenta el criterio de expertos y farmacéuticos componentes del grupo GÉNESIS de la Sociedad Española de Farmacia Hospitalaria, que mostraron una valoración positiva del cuestionario. Este nuevo método de evaluación de análisis por subgrupos se ha añadido al programa Método de Ayuda para la toma de Decisiones y la Realización de Evaluaciones de medicamentos (MADRE) (41), una herramienta estructurada del grupo GÉNESIS que orienta en la selección y posicionamiento terapéutico de los fármacos. También ha sido incorporado por el Grupo Hospitalario de Evaluación de Medicamentos de Andalucía (GHEMA), que evalúa nuevos fármacos y actualiza la Guía Farmacoterapéutica de Hospitales de Andalucía (42).

La herramienta presenta varias limitaciones. Una de ellas es la dificultad de ponderar respuestas en cada uno de los criterios (4). El cuestionario aporta puntuaciones orientativas, útiles para ayudar en el proceso de análisis de datos y decisión clínica. Sin embargo, las puntuaciones no deberían sustituir el buen criterio de los evaluadores debido a la enorme diversidad de escenarios que se presentan en la práctica diaria.

Durante el desarrollo del trabajo, ha surgido la dificultad de establecer un número determinado para delimitar factores que influyen en el análisis por subgrupos, así como el tamaño muestral adecuado de las ramas del estudio y un número de subgrupos pre-especificados. Si bien existen recomendaciones que ayuden a identificar y analizar los factores influyentes en los resultados de los subgrupos (4,8,24), apenas existen valores orientativos. Otro de los obstáculos que encontramos es la escasez de estudios confirmatorios de las hipótesis generadas mediante la interpretación de los análisis por subgrupos. Por ello, la valoración del criterio de consistencia no suele ser una tarea sencilla.

Las principales fortalezas de nuestra herramienta derivan de su estructura. Las cuestiones preliminares garantizan que la decisión tomada según el análisis por subgrupos se fundamente en datos con impacto en la práctica clínica habitual, ya que descarta aquella evidencia científica sin un diseño, variables y factores a estudio adecuados. La toma de decisiones clínicas se ve dificultada por la propia naturaleza del análisis por subgrupos, que incrementa los errores α y β (5). A esto hay que añadir la falta de una metodología e interpretación de la información adecuadamente definidas (8,20). Este cuestionario facilita la valoración ordenada del análisis por subgrupos, guiando al evaluador en el análisis de los criterios adecuados. Una mala interpretación de los resultados de subgrupos puede tener importantes repercusiones clínicas y económicas (43,44). Las puntuaciones de los criterios abordados se han establecido de forma que se evite la aplicación del análisis por subgrupos a la práctica diaria cuando el nivel de incertidumbre sea elevado. Si los factores importantes del análisis por subgrupos no cumplen la metodología definida para considerar fiable un análisis, el sistema de puntuaciones orienta al usuario para descartar el resultado por subgrupos y ceñirse al resultado global del estudio.

Una consideración que se menciona brevemente en el cuestionario es el caso de aquellos ensayos clínicos que engloban poblaciones tan heterogéneas que habrían justificado estudios diferentes, en lugar de precisar un análisis por subgrupos. Este aspecto merece una atención especial por parte de los usuarios, ya que una aplicación estricta del criterio de interacción estadística –valor de $p(i)$ – puede favorecer una

impropia homogeneización de la terapéutica, al ser demasiado cautos para considerar posibles diferencias. Y esto hay que contextualizarlo teniendo en cuenta las citadas limitaciones del análisis por subgrupos. Sin lugar a dudas, la mezcla de poblaciones heterogéneas en los ensayos clínicos que incluyen análisis por subgrupos supone un factor que debe ser abordado profundamente en la investigación terapéutica.

El análisis de subgrupos valora diferencias en el efecto relativo de una intervención en diversas subpoblaciones. Hay ocasiones en las que el diferente riesgo basal de las poblaciones se traduce en diferencias relevantes en el beneficio neto, aunque no existan diferencias absolutas en el efecto (40). Dichas diferencias en el beneficio neto esperable, también pueden afectar al posicionamiento cuando se balancean con los perjuicios y costes. No obstante, esta circunstancia se escapa del ámbito de esta investigación.

Por último, hay que destacar que este tipo de trabajos apoyan la calidad científico-técnica de las decisiones que afectan a la atención sanitaria que reciben los pacientes y promueven la sostenibilidad de los sistemas sanitarios. El análisis por subgrupos adquiere cada vez mayor relevancia debido al aumento del número de publicaciones que lo llevan a cabo, sin que exista una mejora en su interpretación (6,20). Con este estudio, se ha desarrollado y validado un cuestionario útil y fácil de usar, que permite valorar la aplicabilidad de los resultados por subgrupos en la evaluación y posicionamiento de medicamentos.

3.2. Revisión sistemática de la bibliografía disponible sobre análisis por subgrupos de contextos clínicos de alto impacto.

3.2. REVISIÓN SISTEMÁTICA DE LA BIBLIOGRAFÍA DISPONIBLE SOBRE ANÁLISIS POR SUBGRUPOS DE CONTEXTOS CLÍNICOS DE ALTO IMPACTO.

La interpretación de análisis por subgrupos representados en ECA sobre el uso de fármacos en escenarios de elevado impacto clínico-económico supone un tema de interés actual. La valoración de resultados de subpoblaciones puede suscitar controversia (23). Por una parte, las empresas comercializadoras de nuevas moléculas intentan optimizar al máximo su beneficio económico mediante la divulgación de determinados datos de subgrupos (26,31). Por la otra, los sistemas sanitarios se encuentran en una situación comprometida para gestionar los limitados recursos de los que disponen, velando por mantener unos estándares razonables de eficiencia. Asimismo, también existen casos en los que una mala evaluación del análisis por subgrupos puede llegar a restringir nuevas alternativas terapéuticas en determinados subgrupos de población simplemente porque no se tienen en cuenta criterios metodológicos básicos (45).

En este capítulo se mostrarán algunos ejemplos de interpretaciones inadecuadas en los que se ha precisado de una revisión sistemática previa a la aplicación metodológica de la herramienta desarrollada para análisis por subgrupos. En los ejemplos se suele observar que existe cierta heterogeneidad o diferencia estadística entre subpoblaciones [$p(i) < 0.1$] en un ECA, y debido al volumen de evidencia científica disponible, es necesario realizar una evaluación en profundidad que permita dilucidar posibles dudas que surjan en los profesionales sanitarios. Concretamente, los fármacos abordados en este capítulo son daratumumab (46), docetaxel (47) y remdesivir (48).

3.2.1. Esquemas con daratumumab en pacientes con mieloma múltiple de nuevo diagnóstico no candidatos a trasplante que presentan insuficiencia hepática.

Gil-Sierra MD, Briceño-Casado MDP, Fénix-Caballero S, Alegre-Del Rey EJ, Alarcón de la Lastra-Romero C, Sánchez-Hidalgo M. Daratumumab-based therapies in transplant-ineligible patients with untreated multiple myeloma and hepatic dysfunction: A systematic review of subgroup analyses. *J Oncol Pharm Pract.* 2023;29(1):155-161. Doi: 10.1177/10781552211062144.

3.2.1.1. Introducción

El mieloma múltiple (MM) es una neoplasia hematológica con proliferación anormal de células plasmáticas (49). Los pacientes con MM no candidatos a trasplante autólogo de células madre son aquellos que presentan una edad excesiva o determinadas comorbilidades. Esta población puede sufrir efectos secundarios inaceptables asociados a dosis altas de quimioterapia con trasplante de células madre (7). Existe una gran variedad de fármacos para el tratamiento de primera línea de MM en pacientes no candidatos a trasplante, como los inhibidores del proteasoma (50), inmunomoduladores (51) y anticuerpos anti-CD38 (7,45).

Daratumumab es un anticuerpo monoclonal humano IgG1k anti-CD38. El estudio ALCYONE demostró que la combinación de daratumumab con bortezomib, melfalán y prednisona proporciona un considerable aumento en supervivencia global (SG) y supervivencia libre de progresión (SLP) a pacientes con MM no candidatos a trasplante nuevamente diagnosticados (7,52). Asimismo, el ensayo MAIA asoció un incremento de SLP al uso de daratumumab en combinación con lenalidomida y dexametasona frente al esquema lenalidomida-dexametasona en este contexto clínico (45). No obstante, los autores de este ECA comentaron que el regimen de fármacos basado en daratumumab no presentó superioridad frente al brazo control en pacientes con insuficiencia hepática. Esta afirmación fue realizada teniendo en cuenta los datos del análisis por subgrupos del ensayo MAIA (**Figura 4**). Como ya se ha comentado anteriormente, la interpretación del

análisis por subgrupos debe aplicarse con cautela (2,23). Sólo criterios metodológicos estrictos deben ser considerados en la toma de decisiones clínicas para minimizar la incertidumbre asociada al análisis por subgrupos (4).

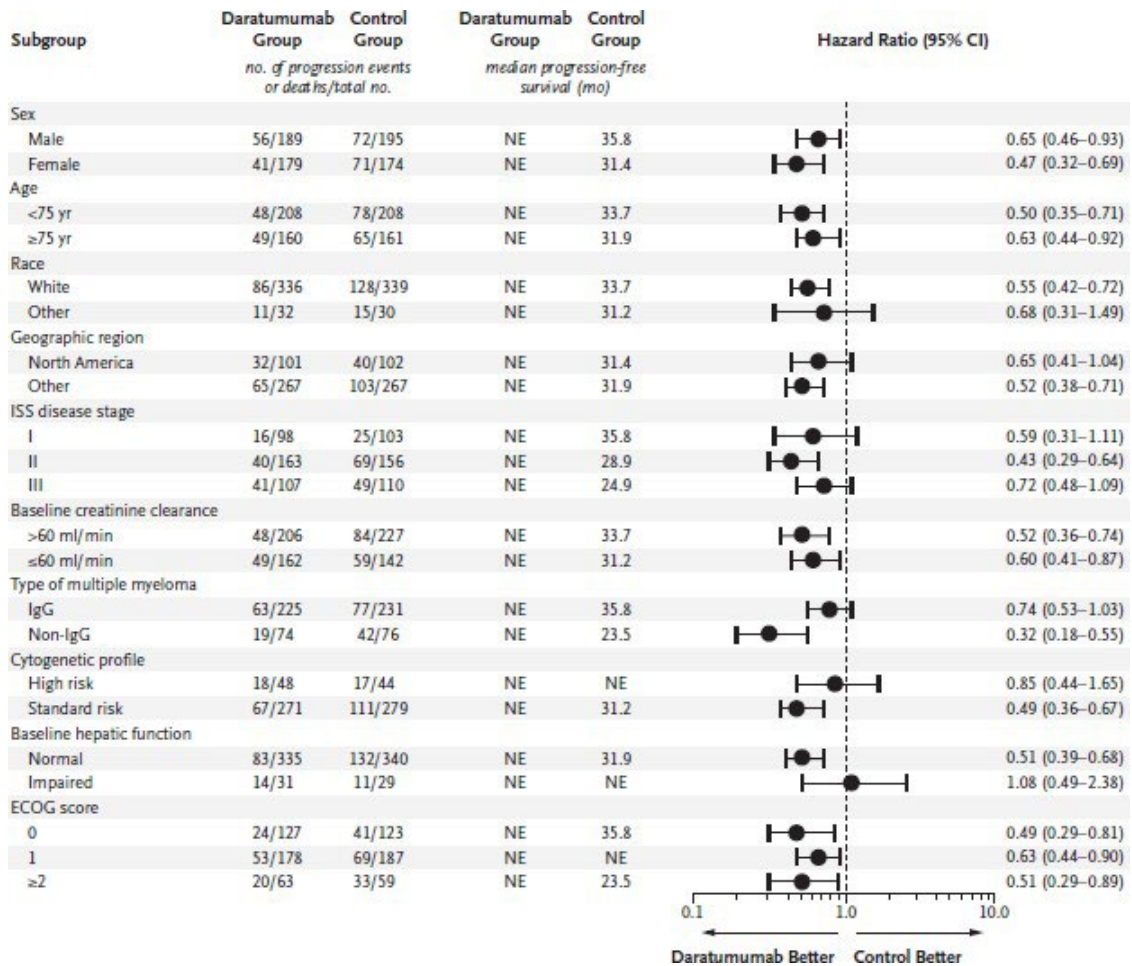


Figura 4. Análisis por subgrupos del estudio MAIA. Tomada de Facon et al. (2019)(45).

La monitorización estrecha de poblaciones con disfunción en algún órgano es esencial para garantizar una adecuada eficacia y seguridad de los tratamientos. Por tanto, el tratamiento de los pacientes con insuficiencia hepática y MM debe seleccionarse cuidadosamente. Teniendo en cuenta el beneficio asociado a los esquemas de daratumumab en primera línea de tratamiento de pacientes con MM no candidatos a trasplante, ¿debería descartarse las terapias con daratumumab para pacientes con daño hepático de acuerdo al análisis por subgrupos del estudio MAIA? A continuación, se desarrolló una revisión sistemática y una valoración metodológica de la eficacia de las

combinaciones de fármacos basadas en daratumumab en pacientes con MM no tratado y daño hepático.

3.2.1.2. Metodología

3.2.1.2.1. Revisión bibliográfica

Se usó el modelo Población-Intervención-Comparador-Resultado-Estudio (criterio PICOS) (53) para abordar este trabajo. Como población se seleccionó a pacientes con MM no tratado previamente y que no son candidatos a trasplante. Los esquemas con daratumumab fueron aceptados como intervención sanitaria. Se permitió cualquier comparador. Los resultados abordados fueron los análisis por subgrupos de SLP y/o SG. El diseño de estudio seleccionado fue ECA fase II o III. Se llevó a cabo una búsqueda sistemática en la base de datos Pubmed® el 11 de octubre de 2020. Se siguieron los criterios de la guía Preferred Reporting Items for a Systematic Review and Meta-analysis (PRISMA) (54). Para desarrollar la búsqueda sistemática, se usó la herramienta de screening “Clinical Queries/Narrow” con la siguiente estrategia de búsqueda: *(Therapy/Narrow[filter]) AND (daratumumab AND myeloma)*.

3.2.1.2.2. Selección de estudios

Durante el proceso de screening, los títulos y resúmenes de los registros obtenidos en la búsqueda fueron revisados para excluir las publicaciones que no cumplían los criterios de inclusión. En cambio, el texto completo de los artículos fue examinado en el proceso de elegibilidad. En este trabajo fueron incluidos ECA fase II/III sobre el uso de daratumumab en pacientes con MM de nuevo diagnóstico que no eran candidatos a trasplante. Estos estudios debían disponer de análisis por subgrupos para SLP y/o SG según daño hepático basal. Se excluyeron estudios con otras características (diseños o poblaciones), pacientes pediátricos (edad menor a 18 años) o aquellos que cumplían los criterios de inclusión en idiomas diferentes al español o inglés.

3.2.1.2.3. Extracción de datos

Se recogieron los siguientes datos de los ECA: autores firmantes, fecha de publicación, tamaño muestral del estudio, seguimiento de los pacientes, intervención sanitaria, brazo comparador, factor diferenciador de subgrupos (existencia de daño hepático basal) y eficacia de los tratamientos analizados.

3.2.1.2.4. Análisis de datos

Se evaluaron los análisis por subgrupos de dos maneras distintas para que el lector pudiera entender mejor los resultados, además de poder comprobar que la herramienta desarrollada en esta Tesis Doctoral aporta unas recomendaciones específicas sobre aplicabilidad de análisis por subgrupos. Por una parte, se valoraron individualmente los criterios de Sun *et al* (4): interacción estadística, pre-especificación de subgrupos, plausibilidad biológica que respalde el efecto observado de la intervención sanitaria, y consistencia de resultados entre estudios. Los valores de $p(i) < 0.1$ fueron aceptados entre subgrupos como diferencia indicativa. Se estimó el valor de $p(i)$ en estudios que no mostraron este dato a partir de valores de hazard ratio e ICs disponibles (12,33).

Posteriormente, se empleó la herramienta validada que se ha desarrollado para valorar la aplicabilidad del análisis por subgrupos (55). Esta herramienta está compuesta por cuestiones preliminares para descartar los análisis por subgrupos sin unas condiciones mínimas relevantes, y un cuestionario. Una respuesta negativa en alguna de las cuestiones preliminares descartó directamente la aplicabilidad del análisis por subgrupos. El cuestionario relacionó una valoración con una puntuación para cada criterio de interpretación del análisis por subgrupos: asociación estadística, plausibilidad biológica y consistencia de resultados. Estas valoraciones y puntuaciones fueron: probable (+3 puntos), posible (+2), dudosa (0) y nula (-3). El sumatorio de las puntuaciones fue relacionado con una recomendación para la aplicabilidad en la práctica clínica: 9-7 puntos (recomendación probable) se asignó con aplicabilidad de resultados de subgrupos; 5-6 puntos (recomendación posible) con aplicabilidad cautelosa ante dificultad de uso, escasa tolerancia o coste excesivo de alternativas

terapéuticas; 3-4 puntos (recomendación dudosa) con evitar aplicabilidad de datos de subgrupos salvo excepciones; y <3 puntos (recomendación nula) con no aplicación. Si la asociación estadística o consistencia fueron nulas, la aplicabilidad de los resultados de subgrupos se descartó directamente y se llevó a cabo una “recomendación nula”.

3.2.1.2.5. Riesgo de sesgo

Se consideraron características basales de los pacientes con influencia en los resultados: edad, estado funcional según Eastern Cooperative Oncology Group (ECOG), estadio de la enfermedad según ISS y perfil citogenético (56). Los porcentajes fueron calculados usando las poblaciones globales de los estudios.

3.2.1.3. Resultados

3.2.1.3.1. Búsqueda de la literatura y características de los estudios

Se obtuvieron 25 resultados en la búsqueda sistemática. De ellos, se excluyeron 10 registros por tratarse de estudios con diseño diferente a ECA fase II/III, 9 resultados por analizar un contexto clínico diferente, 2 registros por no evaluar el fármaco daratumumab y un resultado por no presentar análisis por subgrupos. Por tanto, 3 estudios fueron incluidos: 2 ECA y una actualización de uno de ellos. La búsqueda sistemática de la literatura es detallada en la **Figura 5**. Los datos de los estudios seleccionados son mostrados en la **Tabla 2**.

Figura 1. Revisión sistemática de la literatura

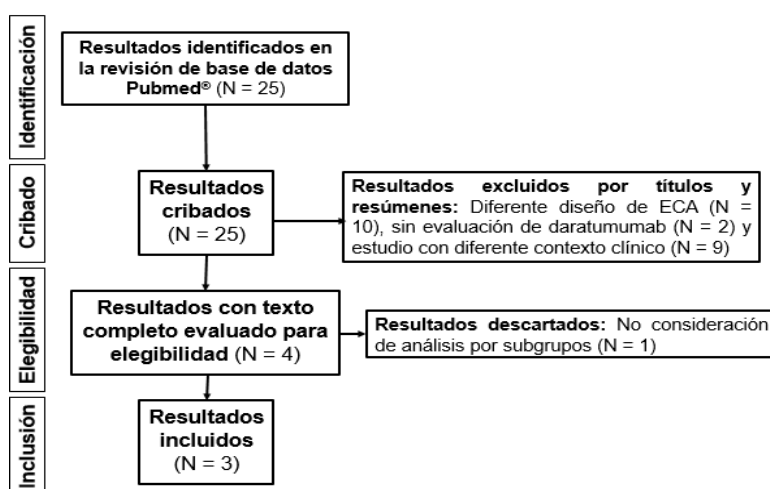


Figura 5. Revisión sistemática de la literatura

Tabla 2. Datos de los ensayos clínicos incluidos

Autores	Fecha de publicación	Tamaño muestral del ensayo (nº de pacientes)	Seguimiento (mediana, meses)	Intervención	Comparador	Factor diferenciador de subgrupos (función hepática basal alterada)	Tamaño muestral del subgrupo con función hepática basal normal† (nº pacientes)	Tamaño muestral del subgrupo con función hepática basal alterada (nº pacientes)	HR para variables (95%IC) en población global‡	HR para SLP (95%IC) en subgrupo con función hepática normal‡	HR para variables (95%IC) en subgrupos con daño hepático‡
Mateos <i>et al</i> (Ensayo ALCYONE)	8 Febrero 2018	706	16,5	Daratumumab, bortezomib, melfalan y prednisona	Bortezomib, melfalan y prednisona	Nivel de bilirrubina total \leq límite superior del rango normal (ULN) y nivel de aspartato aminotransferasa $>$ ULN, o nivel de bilirrubina total $>$ ULN	604	98	SLP: 0,50 (0,38-0,65)	SLP: 0,53 (0,40-0,71)	SLP: 0,42 (0,22-0,80)
Facon <i>et al</i> (ensayo MAIA)	30 Mayo 2019	737	28	Daratumumab, lenalidomida y dexametasona	Lenalidomida y dexametasona	Nivel de bilirrubina total \leq límite superior del rango normal (ULN) y nivel de aspartato aminotransferasa $>$ ULN, o nivel de bilirrubina total $>$ ULN	675	60	SLP: 0,56 (0,43-0,73)	SLP: 0,51 (0,39-0,68)	SLP: 1,08 (0,49-2,38)
Mateos <i>et al</i> (actualización del ensayo ALCYONE)	11 Enero 2020	706	40,1	Daratumumab, bortezomib, melfalan y prednisona	Bortezomib, melfalan y prednisona	Nivel de bilirrubina total \leq límite superior del rango normal (ULN) y nivel de aspartato aminotransferasa $>$ ULN, o nivel de bilirrubina total $>$ ULN	604	98	SG: 0,60 (0,46-0,80)	SG: 0,66 (0,48-0,89)	SG: 0,47 (0,24-0,90)

†El subgrupo con función hepática normal es el grupo de pacientes que no cumple los criterios que definen al subgrupo con función hepática alterada. ‡HR = *hazard ratio*. Las variables fueron SLP (supervivencia libre de progresión) y SG (supervivencia global). %IC= porcentaje de intervalo de confianza.

3.2.1.3.2. Análisis por subgrupos

El ensayo ALCYONE fue publicado el 8 de febrero de 2018, incluyendo un análisis por subgrupos para SLP (7). Primero se aplicaron los criterios sobre análisis por subgrupos de Sun *et al* (4). $P(i)$ no fue descrita en el ensayo ALCYONE y tuvo que ser estimada. No se encontró heterogeneidad entre los subgrupos según la función hepática [$p(i) = 0,582$]. La pre-especificación de subgrupos fue definida en el ECA. Con respecto a la plausibilidad biológica, en un estudio publicado previamente se observó una exposición similar de daratumumab en pacientes con MM que tenían función hepática normal y aquellos con daño hepático (57). Asimismo, este estudio no encontró diferencia de eficacia entre estos dos subgrupos. Teniendo en cuenta esto, parece razonable afirmar que no hubo datos consistentes de diferencia de eficacia para daratumumab entre subgrupos según función hepática hasta la publicación del ensayo ALCYONE.

Posteriormente, se usó la herramienta validada que se ha desarrollado sobre aplicabilidad de análisis por subgrupos (55). Se descartó directamente la aplicabilidad del análisis por subgrupos en el estudio ALCYONE por respuesta negativa de una cuestión preliminar sobre existencia de diferencia de efecto entre los subgrupos [$p(i) > 0.1$].

El ensayo MAIA fue publicado el 30 de mayo de 2019 con un análisis por subgrupos para SLP (45). De nuevo, los criterios sobre interpretación de análisis por subgrupos de Sun *et al* fueron considerados (4). El ensayo MAIA no mostró datos sobre $p(i)$, por lo que tuvo que ser calculado. Se observó heterogeneidad entre subgrupos según función hepática, al estimar una diferencia indicativa de $p(i) = 0,0792$. Este análisis por subgrupos fue pre-especificado en el ECA. En lo que se refiere a la plausibilidad biológica, la literatura previa no había demostrado datos que respaldaran que una especial exposición y eficacia de daratumumab en población con daño hepático fuera cierta (57). Por otra parte, no se encontró consistencia de datos entre la aparente diferencia en subgrupos según función hepática observada en el ensayo MAIA con los resultados descritos en el ensayo ALCYONE (sin heterogeneidad entre subgrupos, $p(i) > 0.1$).

Después de esto, se aplicó la herramienta que hemos desarrollado sobre análisis por subgrupos (55). En el ensayo MAIA, todas las cuestiones preliminares tuvieron respuesta positiva. Los resultados del cuestionario fueron: asociación estadística con valoración “nula” (-3 puntos), plausibilidad biológica “dudosa” (0 puntos) y consistencia “nula” (-3 puntos). Por tanto, la aplicabilidad del análisis por subgrupos se descartó directamente por presentar asociación estadística y consistencia “nulas” (recomendación final nula).

El estudio de actualización de datos del ensayo ALCYONE fue publicado el 11 de enero de 2020 con un análisis por subgrupos para SG (52). Primero, se aplicaron los criterios de valoración de subgrupos de Sun *et al* (4). Esta actualización del ensayo ALCYONE no mostró valores de $p(i)$. Por tanto, se calculó y no se encontró heterogeneidad [$p(i) = 0.362$]. Los subgrupos según función hepática se encontraban pre-especificados. En cuanto a la plausibilidad biológica, la hipótesis de una especial exposición y eficacia de daratumumab en una población con insuficiencia hepática no fue respaldada por bibliografía previa (57). A diferencia de análisis por subgrupos realizados en función de SLP en estudios previos (7,45), la actualización de datos del ensayo ALCYONE incorporó un análisis por subgrupos para OS. No obstante, estas dos variables podrían estar relacionadas. Una mayor SLP podría traducirse en un incremento de SG para los pacientes en este contexto clínico. Por otra parte, existe inconsistencia entre la heterogeneidad observada en subgrupos para SLP del ensayo MAIA y la ausencia de interacción estadística para SLP y SG del estudio ALCYONE.

Por último, usamos nuestra herramienta validada de aplicabilidad de resultados por subgrupos (55). En este caso, la cuestión preliminar sobre la existencia de diferencia de efecto entre los subgrupos obtuvo respuesta negativa [$p(i) > 0.1$]. Por tanto, se descartó directamente la aplicabilidad del análisis por subgrupos. La **Tabla 3** representa un resumen de la valoración del análisis por subgrupos de los estudios incluidos.

Tabla 3. Resumen de evaluación sobre análisis por subgrupos

Metodología	Criterios		Mateos <i>et al</i> (ensayo ALCYONE), evaluación de SLP†	Facon <i>et al</i> (ensayo MAIA), evaluación de SLP†	Mateos <i>et al</i> (actualización de resultados del estudio ALCYONE), evaluación de SG‡
Sun <i>et al.</i>	Interacción estadística		No [p(i) = 0.582]	Sí, interacción estadística indicativa [p(i) = 0.0792]	No [p(i) = 0.362]
	Pre-especificación de análisis por subgrupos		Sí	Sí	Sí
	Plausibilidad biológica		No	No	No
	Consistencia de resultados de subgrupos		No	No	No
Herramienta validada (Gil-Sierra <i>et al.</i>)	Cuestiones preliminares	Nivel más alto de evidencia del estudio con análisis por subgrupos	Sí	Sí	Sí
		Relevancia clínica clara de variable considerada o variable primaria subrogada	Sí	Sí	Sí
		Existencia de diferencia en el efecto entre los subgrupos para el factor analizado [p(i) < 0.1]	No	Sí	No
		Factor determinante del análisis por subgrupos previo a la intervención sanitaria	No aplicado	Sí	No aplicado
	Cuestionario		No aplicado	Aplicado	No aplicado
Asociación estadística (puntuación)		No aplicado	Nulo (-3 puntos)	No aplicado	
Plausibilidad biológica (puntuación)		No aplicado	Dudoso (0 puntos)	No aplicado	
Consistencia (puntuación)		No aplicado	Nulo (-3 puntos)	No aplicado	
Recomendación		No aplicado	Nulo: Análisis por subgrupos no debe aplicarse	No aplicado	

†SLP: supervivencia libre de progresión. ‡SG: supervivencia global.

3.2.1.3.3. Riesgo de sesgo

La edad media de los pacientes (brazo con intervención-brazo con control) fue 71-71 años en el ensayo ALCYONE y 73-74 años en estudio MAIA (7,45). La proporción de pacientes con ≥ 75 años fue 29.9% en ALCYONE y 43.5% en MAIA. La distribución del estado funcional según ECOG en el estudio ALCYONE fue la siguiente: 50.3% de pacientes con ECOG 1 y 24.6% con ECOG 2. La distribución del estado funcional según ECOG en el ensayo MAIA se correspondió con un 49.5% de pacientes con ECOG 1 y 16.5% con ECOG 2. El estudio ALCYONE incluyó un 42.3% de pacientes con ISS II y un 38.4% con ISS III. El ensayo MAIA reclutó un 43.3% de pacientes con ISS II y un 29.4% con ISS III. Las proporciones de pacientes con riesgo citogenético alto fueron: 13.9% en el ensayo ALCYONE y 14.3% en el estudio MAIA.

3.2.1.4. Discusión

En este trabajo no se observaron diferencias consistentes de eficacia de esquemas basados en daratumumab entre la subpoblación con daño hepático y la que presenta función hepática normal en MM de nuevo diagnóstico sin posibilidad de trasplante. Los resultados obtenidos en nuestra herramienta validada sobre interpretación y aplicabilidad del análisis por subgrupos recomiendan no considerar heterogeneidad entre los subgrupos de pacientes según función hepática.

La población del ensayo ALCYONE presentó menor edad que el estudio MAIA. Casi la mitad de los pacientes reclutados en el ensayo MAIA tenían igual o más de 75 años. Sin embargo, el estudio ALCYONE presentó una mayor proporción de pacientes con peor estado funcional según ECOG y estadio de la enfermedad según ISS. Algunas limitaciones del presente estudio es que no existían datos actualizados sobre SG de daratumumab en combinación con lenalidomida y dexametasona cuando este trabajo fue desarrollado, además del uso de diferentes comparadores en los ECA y el uso de medicación concomitante dispar en las combinaciones de tratamiento que incluían daratumumab.

Los esquemas basados en daratumumab han demostrado incrementos de SLP y SG en primera línea de tratamiento de pacientes con MM (7,45,52). SLP es una variable de carácter irreversible que puede influir positivamente en el retraso del uso de líneas de tratamientos posteriores y también en el aumento de SG de pacientes hematológicos con etapas tempranas de enfermedad. Esto reviste una gran importancia, ya que SG y la calidad de vida de los pacientes son las variables más relevantes en onco-hematología (58–61). Limitar el uso de los regímenes de fármacos que contienen daratumumab en pacientes no candidatos a trasplante con MM no tratado y daño renal podría reducir su supervivencia. Considerando que tampoco se encontró una hipótesis razonable –y consistente en los diferentes esquemas– que justificara la pérdida del beneficio obtenido de la administración de daratumumab en población con disfunción hepática, es desaconsejable aplicar un análisis por subgrupos con tan elevada incertidumbre en estos pacientes. No hay que olvidar las limitaciones del análisis por subgrupos, ya que interpretaciones inadecuadas pueden llevar a la toma de decisiones clínicas con repercusiones negativas para los pacientes.

La gran variedad de alternativas terapéuticas en MM y de perfiles diferentes de pacientes requiere el uso de herramientas validadas y criterios estandarizados para un adecuado posicionamiento terapéutico de los fármacos. Una fortaleza de nuestro trabajo es la aplicación de metodologías sistemáticas, lo cual es especialmente importante ante entidades tan controvertidas como es el análisis por subgrupos. Esta revisión sistemática no encontró suficiente evidencia científica que respalde una diferencia de eficacia de las combinaciones de fármacos basados en daratumumab entre pacientes con MM de nuevo diagnóstico con daño hepático y aquellos con función hepática normal. Por tanto, y siempre que se garantice la seguridad de los pacientes, no estaría justificada una restricción del uso de los esquemas de daratumumab en pacientes con disfunción hepática.

3.2.2. Docetaxel en cáncer de próstata metastásico hormonosensible con alta y baja carga tumoral.

Gil-Sierra MD, Alegre-Del Rey EJ, Alarcon de la Lastra-Romero C, Sánchez-Hidalgo M. Use of docetaxel in low- and high-burden metastatic hormone-sensitive prostate cancer: A systematic review and assessment of subgroup analyses. *J Oncol Pharm Pract.* 2021;27(7):1743-1750. Doi: 10.1177/10781552211037322.

3.2.2.1. Introducción

El cáncer de próstata es la neoplasia más frecuente en varones (62,63). Se ha estimado que se diagnosticaron más de un millón de nuevos casos de cáncer de próstata en 2018, asociándose más de 300.000 muertes. Los pacientes con cáncer de próstata metastásico hormonosensible (mHSPC) son aquellos que presentan enfermedad metastásica y que no han sido tratados con terapia hormonal, o que presentan respuesta a este tratamiento.

La quimioterapia con docetaxel en combinación con terapia de deprivación androgénica (ADT, del inglés *androgen deprivation therapy*) ha mostrado importantes beneficios clínicos en mHSPC (64). No obstante, existe bibliografía que se ha usado para respaldar un efecto diferente de docetaxel en mHSPC según la carga de la enfermedad (64,65) (**Figura 6**). Estos estudios sugieren un aumento de SG en pacientes con alto volumen de la enfermedad, y ausencia de beneficio en pacientes con bajo volumen metastásico. Por otra parte, otro ensayo ha demostrado una mejora en SG asociada a docetaxel para pacientes con mHSPC, independientemente de la carga de la enfermedad (66). Por tanto, se puede comprobar que el uso de docetaxel en combinación con ADT en mHSPC presenta una considerable controversia. A todo esto, hay que sumar que existen otras alternativas terapéuticas como abiraterona que aportan un gran beneficio en este contexto clínico (67). Teniendo en cuenta todo lo anterior, el posicionamiento terapéutico de las diferentes terapias y la toma de decisiones clínicas en mHSPC –sobre todo en las diferentes subpoblaciones según el volumen de la enfermedad– es un desafío para los profesionales sanitarios.

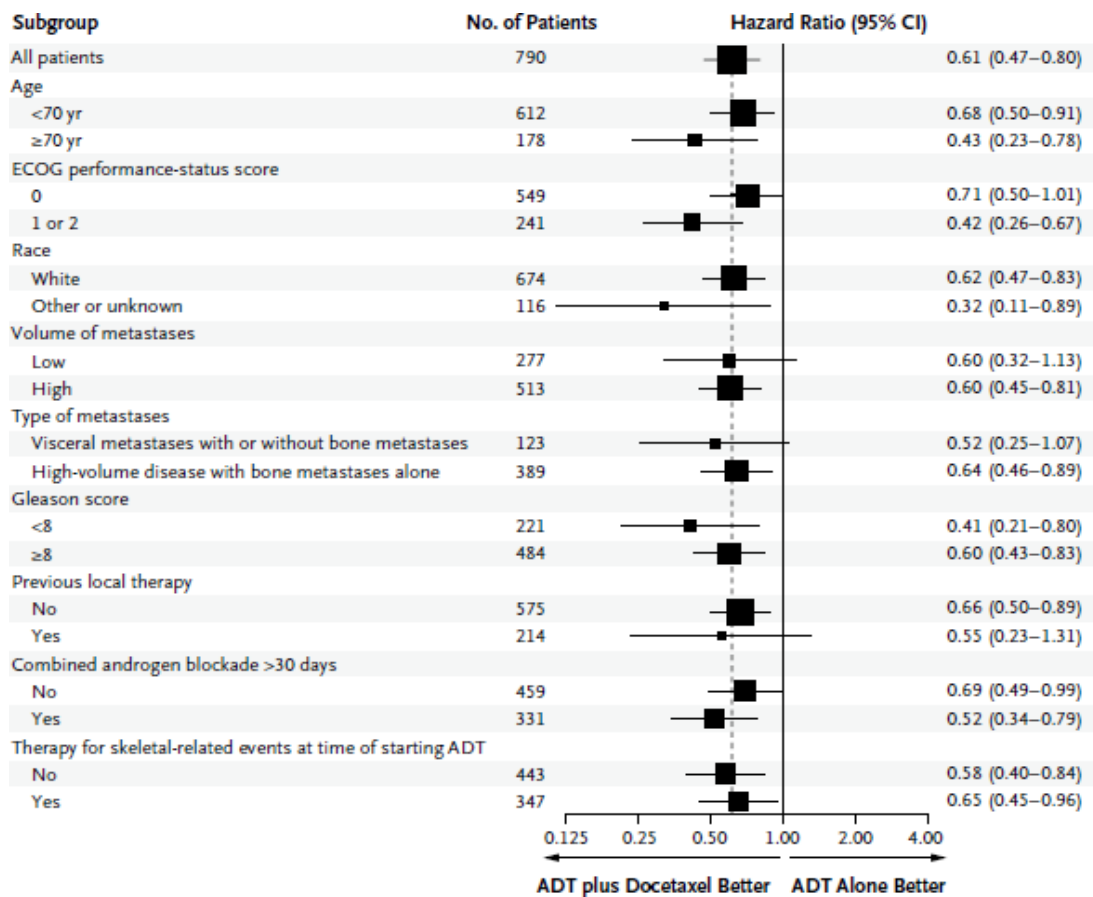


Figura 6. Análisis por subgrupos del estudio CHARTED. Tomada de Sweeney et al. (2015) (64)

Algunas de las afirmaciones sobre la eficacia de docetaxel en los diferentes grupos de pacientes se han llevado a cabo teniendo en cuenta los datos de análisis por subgrupos, los cuales deben ser interpretados siguiendo criterios metodológicos sólidos (55). Ante este escenario, ¿qué evidencia científica respalda el uso de docetaxel para pacientes con mHSPC independientemente del volumen metastásico? ¿Qué datos sugieren la ausencia de beneficio de docetaxel en pacientes con mHSPC y baja carga de la enfermedad? A continuación, vamos a analizar la hipotética diferencia de eficacia de docetaxel en los subgrupos de alto y bajo volumen metastásico de mHSPC, realizando una revisión sistemática de la literatura y aplicando metodologías con criterios objetivos.

3.2.2.2. Metodología

3.2.2.2.1. Revisión bibliográfica

La pregunta de investigación se formuló teniendo en cuenta la estrategia PICOS (53), que considera la población, intervención, comparador, resultado y diseño de los estudios. La población incluyó pacientes con mHSPC. La intervención seleccionada fue la terapia con docetaxel. Se incluyó cualquier comparador. El resultado fueron los datos de SG del análisis por subgrupos según el volumen tumoral. Con respecto al diseño de los estudios, se seleccionaron ECA y estudios con análisis por subgrupos sobre datos de ECA previos. Se realizó una búsqueda sistemática en la literatura de acuerdo con las directrices PRISMA (54). Esta revisión se desarrolló el 25 de septiembre 2020 en la base de datos Pubmed®. Se utilizó la herramienta 'Clinical Queries/Narrow' con la siguiente estrategia de búsqueda: *[(Therapy/Narrow [filter]) AND (docetaxel AND prostate cancer AND hormone sensitive)]*. Posteriormente, se realizó un rastreo de citas.

3.2.2.2.2. Selección de estudios

En el proceso de cribado, los títulos y resúmenes obtenidos en la revisión se examinaron para identificar y excluir publicaciones que no cumplieran los criterios de inclusión predefinidos. En el proceso de elegibilidad, se revisó el texto completo de los resultados. En este trabajo se incluyeron los estudios sobre el uso de docetaxel en pacientes con mHSPC y un análisis de subgrupos con datos de SG según el volumen metastásico. El subgrupo de alto volumen tumoral fue definido como aquellos pacientes con presencia de metástasis viscerales o cuatro o más lesiones óseas, con al menos una en cuerpos vertebrales y pelvis. El subgrupo de pacientes con bajo volumen tumoral incluyó pacientes que no cumplieron con estos criterios. Los resultados obtenidos en la revisión en un idioma distinto del español o inglés fueron excluidos.

3.2.2.2.3. Extracción de datos

Los estudios fueron evaluados por fecha de publicación y se registraron los siguientes datos: autores, fecha de publicación, tamaño muestral, duración del seguimiento de los pacientes, brazo de intervención, brazo de comparador y factores utilizados para diferenciar los subgrupos.

3.2.2.2.4. Análisis de datos

Se evaluaron los análisis por subgrupos de dos maneras distintas para que el lector pudiera comprender mejor los resultados, además de poder comprobar que la herramienta desarrollada en esta tesis doctoral aporta unas recomendaciones específicas sobre aplicabilidad de análisis por subgrupos. Primero, los análisis por subgrupos fueron evaluados según los criterios descritos en Sun *et al* (4): interacción estadística, pre-especificación de subgrupos, plausibilidad biológica que respalde al efecto de la intervención sanitaria en los subgrupos, y consistencia de los resultados dentro del mismo estudio (consistencia interna) o en comparación con los datos de otros estudios. Valores de $p(i) < 0.1$ son aceptados entre subgrupos como una diferencia indicativa debido a la disminución de la potencia estadística de los análisis por subgrupos (2). Se estimó el valor de $p(i)$ en los estudios sin este dato mediante calculadoras con valores de riesgo relativo, odds ratio, hazard ratio e ICs (12,68–72).

Posteriormente, se usó nuestra herramienta validada para valorar la aplicabilidad del análisis por subgrupos (55). Esta herramienta se divide en dos partes: cuestiones preliminares para descartar los análisis por subgrupos sin condiciones mínimas relevantes, y un *checklist*. Las cuestiones preliminares consideran el nivel de evidencia del estudio, la relevancia clínica de la variable evaluada, la diferencia de efecto entre los subgrupos [$p(i) < 0.1$] y la secuencia temporal entre el factor diferenciador de subpoblaciones y la intervención sanitaria. La contestación de todas las cuestiones preliminares, desarrollo del *checklist* y aplicabilidad del análisis por subgrupos son descartados ante una respuesta negativa en alguna de las cuestiones preliminares. El *checklist* asocia valoraciones con puntuaciones para cada criterio de la interpretación

del análisis por subgrupos: asociación estadística, que incluye $p(i)$, pre-especificación del análisis, tamaño muestral de subgrupos, número de factores evaluados y resultado global del estudio; plausibilidad biológica y consistencia de resultados. Los rangos de las valoraciones y puntuaciones son: probable (+3 puntos), posible (+2), dudosa (0) y nula (-3). El sumatorio de las puntuaciones de los tres criterios (asociación estadística, plausibilidad biológica y consistencia) es relacionada con una recomendación para la aplicabilidad en la práctica clínica de los resultados de los subgrupos. Puntuaciones más altas del checklist se corresponden con una mayor aplicabilidad de los resultados de los subgrupos: 9-7 puntos (recomendación probable) se corresponde con aplicabilidad del análisis por subgrupos; 5-6 puntos (recomendación posible) con aplicabilidad prudente ante escasa tolerancia, dificultad de uso o coste de otras intervenciones; 3-4 puntos (recomendación dudosa) se relaciona con evitar aplicabilidad salvo excepciones; y <3 puntos (recomendación nula) se relaciona con la no aplicación del análisis por subgrupos. Si la asociación estadística o consistencia fueran consideradas nulas, la aplicabilidad del análisis se descartaría directamente sin considerar el sumatorio de los criterios (recomendación nula).

3.2.2.2.5. Riesgo de sesgo

El pronóstico de los pacientes de un ECA puede influir en los resultados obtenidos por una intervención sanitaria. En este caso, se recogieron los siguientes datos de las poblaciones estudiadas: puntuación de escala Gleason, niveles séricos de antígeno prostático específico (PSA, del inglés *prostate-specific antigen*), estado funcional según ECOG, estado y tratamiento previo para el cáncer de próstata. Asimismo, también se revisaron los tamaños muestrales de los subgrupos.

3.2.2.3. Resultados

3.2.2.3.1. Búsqueda de la literatura y características de los estudios

Se encontraron 31 registros en la búsqueda sistemática de la base de datos Pubmed®. Durante el rastreo de citas, se encontró otro ECA que evaluó el efecto de docetaxel. Del total de los 32 registros, se excluyeron 9 resultados de búsqueda por no presentar diseño de ECA, 7 por evaluar a otro fármaco diferente a docetaxel, 6 por no medir el efecto de un fármaco, 5 por no incluir análisis por subgrupos según volumen metastásico y 1 por no valorar la SG. El ECA identificado en el rastreo de citas no fue seleccionado por ausencia de análisis por subgrupos según carga de la enfermedad metastásica (73). Por tanto, cuatro ECA fueron incluidos en este estudio. La **Figura 7** muestra el proceso de búsqueda sistemática de la literatura según el protocolo PRISMA y la **Tabla 4** resume los datos de los estudios seleccionados.

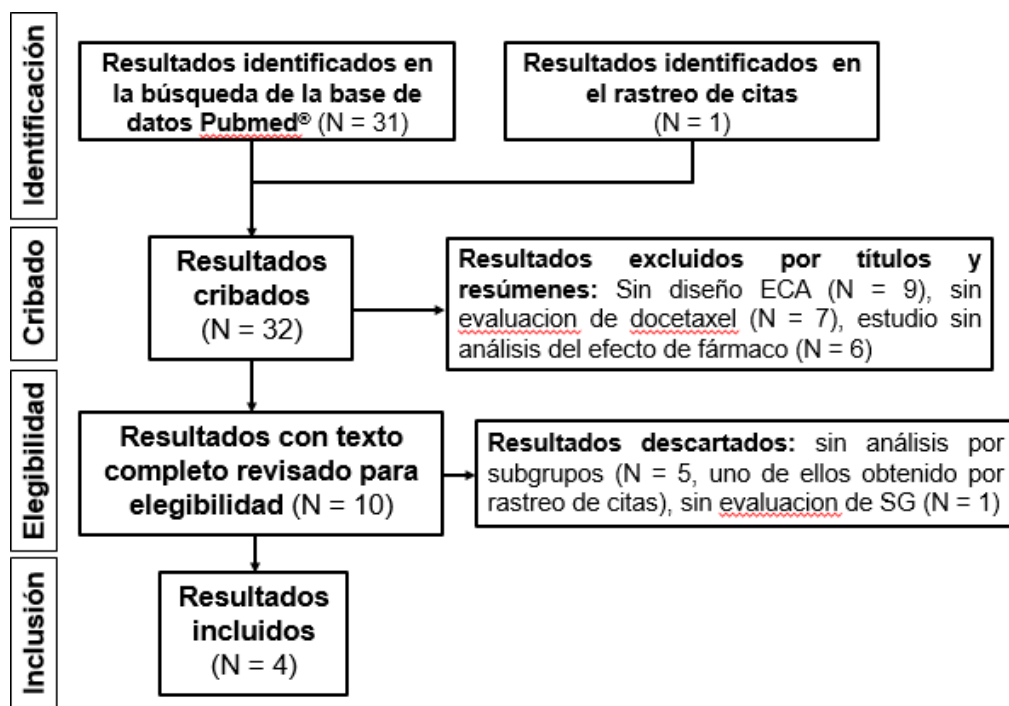


Figura 7. Revisión sistemática de la literatura.

Tabla 4. Datos de los ECA seleccionados

Autores	Fecha de publicación	Tamaño muestral del ensayo (nº de pacientes)	Seguimiento (mediana, meses)	Intervención	Comparador	Factor diferenciador de subgrupos (localización y número de metástasis en volumen elevado de metástasis)	Tamaño muestral del subgrupo de bajo volumen de metástasis[†] (nº pacientes)	Tamaño muestral del subgrupo de elevado volumen de metástasis (nº pacientes)
Sweeney <i>et al</i> (Estudio CHARTED)	5 agosto 2015	790	28,9	ADT + Docetaxel 75 mg/m ² cada 3 semanas durante 6 ciclos	ADT	Presencia de metástasis viscerales o ≥4 lesiones óseas con ≥1 en cuerpos vertebrales y pelvis	277	513
Gravis <i>et al</i> (análisis <i>post hoc</i> de estudio GETUG-AFU15)	21 noviembre 2015	385	83,9	ADT + Docetaxel 75 mg/m ² cada 3 semanas durante 9 ciclos	ADT	Presencia de metástasis viscerales o ≥4 lesiones óseas con ≥1 en cuerpos vertebrales y pelvis	202	183
Kyriakopoulos <i>et al</i> (actualización del estudio CHARTED)	31 enero 2018	790	53,7	ADT + Docetaxel 75 mg/m ² cada 3 semanas durante 6 ciclos	ADT	Presencia de metástasis viscerales o ≥4 lesiones óseas con ≥1 en cuerpos vertebrales y pelvis	277	513
Clarke <i>et al</i> (actualización del estudio STAMPEDE)	27 septiembre 2019	1086	78,2	ADT + Docetaxel 75 mg/m ² cada 3 semanas durante 6 ciclos	ADT	Presencia de metástasis viscerales o ≥4 lesiones óseas con ≥1 en cuerpos vertebrales y pelvis	362	468

ADT: terapia de deprivación androgénica. †El subgrupo de pacientes con bajo volumen de metástasis se corresponde con el grupo de pacientes que no cumple los criterios que definen al subgrupo de alto volumen de metástasis.

3.2.2.3.2. Análisis por subgrupos

El ensayo clínico de Sweeney *et al* fue publicado online el 5 de agosto de 2015 (64). A este trabajo se le aplicaron los criterios para la interpretación de análisis por subgrupos de Sun *et al* (4). La $p(i)$ para SG no fue descrita en este estudio. Por tanto, $p(i)$ fue calculada y no se encontró heterogeneidad entre los subgrupos de alto y bajo volumen metastásico [$p(i) = 1$]. La pre-especificación de los subgrupos según la extensión de la enfermedad no se encuentra definida en este ECA. No obstante, los pacientes fueron estratificados según este factor. La plausibilidad biológica que respaldaría una diferencia de efecto en los subgrupos de alto y bajo volumen metastásico podría parecer razonable, ya que se puede llegar a pensar que los pacientes con bajo volumen tienen un mejor pronóstico de la enfermedad por una menor extensión neoplásica. Además, un abordaje terapéutico más agresivo podría obtener un efecto diferencial en la población con peor pronóstico. Esto podría basarse en resultados de subgrupos de estudios previos de otros fármacos, en los que se observaron aparentes diferencias entre las medianas de SG de los pacientes con enfermedad extendida y mínima, así como la interpretación de gráficas (74,75). Estos ECA no aportaron suficientes datos para evaluar la heterogeneidad entre los subgrupos abordados, por lo que no se pudo determinar si existía o no consistencia de los resultados con Sweeney *et al*. Sin embargo, este último estudio demostró cierta consistencia interna, ya que en otros subgrupos que podrían ser predictores de mala evolución de la enfermedad, como la escala Gleason, no se observó diferencia de efecto entre las subpoblaciones. Asimismo, el estudio GETUG-AFU 15 no mostró heterogeneidad entre subgrupos de diferente pronóstico: alto, intermedio y bajo riesgo. Esta clasificación de diferentes pronósticos basada en localización de enfermedad, estado del paciente, niveles del PSA, y escala Gleason (76,77).

Posteriormente, se usó nuestra herramienta validada de aplicabilidad clínica para análisis por subgrupos (55). No se contestó positivamente a la cuestión preliminar sobre existencia de diferencia de efecto entre los subgrupos [$p(i) > 0.1$]. Por tanto, no se continuó con la aplicación de la herramienta y se descartó directamente el uso de los

datos de los subgrupos. La **Tabla 5** muestra un resumen de la interpretación del análisis por subgrupos descrita para el ensayo de Sweeney *et al* (64).

El estudio de Gravis *et al* fue publicado online el 21 de noviembre de 2015 (78). Se trata de un trabajo que retrospectivamente evaluó los datos del ensayo GETUG-AFU 15 (77), llevando a cabo análisis de subgrupos *post hoc* según el volumen metastásico. El resultado global del estudio fue negativo, es decir, no se obtuvo diferencia estadísticamente significativa para SG entre los pacientes tratados con docetaxel + ADT y los que recibieron ADT. Nuevamente, se usaron los criterios de valoración de análisis por subgrupos de Sun *et al* (4). Este ensayo presentó el valor de $p(i)$ para SG [$p(i) = 0.4$], no observándose heterogeneidad entre los subgrupos. Debido al carácter *post hoc* del estudio, no puede aceptarse que haya pre-especificación del análisis por subgrupos según la extensión de la enfermedad. La plausibilidad biológica de que existiera diferencia entre los subgrupos pudiera parecer razonable, ya que se podría pensar que los pacientes que tienen una mayor carga tumoral pueden verse beneficiados en mayor medida por una terapia, a priori, más agresiva como docetaxel. A esto habría que sumar los aparentes resultados de los estudios previos ya mencionados en otros fármacos (74,75). Se observa consistencia interna –de ausencia de heterogeneidad entre subgrupos– para otras variables del mismo estudio: SLP de PSA, radiográfica o muerte [$p(i) = 0.1840$] y progresión radiográfica o muerte [$p(i) = 0.2568$]. Sin embargo, el resultado global negativo de Gravis *et al* genera inconsistencia con el beneficio en SG obtenido por el uso de docetaxel + ADT en Sweeney *et al* (64).

Después, se volvió a emplear nuestra herramienta validada de aplicabilidad clínica para el análisis por subgrupos (55). Se contestó negativamente a la cuestión preliminar sobre la existencia de diferencia de efecto entre los subgrupos según volumen de metástasis [$p(i) > 0.1$], por lo que no se continuó aplicando la herramienta. Por tanto, se descartó directamente la aplicabilidad del análisis por subgrupos. La **Tabla 5** ilustra un resumen de la interpretación del análisis por subgrupos del estudio de Gravis *et al* (78).

El estudio de Kyriakopoulos *et al* fue publicado online el 31 de enero de 2018 (65). Se trata de un ensayo en el que se actualizaron los resultados del estudio CHARTED (64). Primero, aplicamos los criterios para la interpretación de análisis por subgrupos de Sun *et al* (4). No se detalló el valor de $p(i)$ en Kyriakopoulos *et al*. Se encontró heterogeneidad [$p(i) = 0.03$] con los datos proveídos por el artículo. El análisis por subgrupos según el volumen de metástasis fue pre-especificado e incluido en la estratificación de los pacientes. La plausibilidad biológica sobre una diferencia de efecto entre subgrupos de alta y baja carga de enfermedad se basó en posibles diferencias biológicas por la extensión de la enfermedad. Se sugirió que un uso temprano de docetaxel podrían influir favorablemente en resultados de tratamientos posteriores en pacientes con alto volumen de metástasis. Además, es importante recordar a los resultados aparentes ya mencionados de estudios anteriores sobre otros fármacos -como flutamida- o Gravis *et al* (74,75,78). Con respecto a la consistencia, se descartó la comparación de datos de Kyriakopoulos *et al* con resultados menos maduros del propio estudio CHARTED (64). Por ello, sólo quedaría comparar los resultados del análisis por subgrupos de este estudio con los obtenidos por docetaxel + ADT en Gravis *et al* (78). No se observó consistencia entre los resultados de ambos estudios. En Kyriakopoulos *et al* se encontró diferencia de efecto entre los subgrupos con beneficio para los pacientes de mayor carga tumoral –sin beneficio para el de menor volumen de metástasis–. Sin embargo, en Gravis *et al* no se encontró heterogeneidad ni beneficio para el global de los pacientes.

Posteriormente, se utilizó nuestra herramienta validada de aplicabilidad clínica para el análisis por subgrupos (55). Las cuatro cuestiones preliminares fueron respondidas positivamente. Al aplicar el checklist, la asociación estadística presentó una valoración “probable” (+3 puntos). La plausibilidad biológica también obtuvo una valoración de “probable” (+3 puntos) y la consistencia fue “nula” (-3 puntos). La aplicabilidad del análisis se descartó directamente sin considerar el sumatorio por presentar una consistencia “nula” (recomendación final nula). Un resumen de la interpretación del análisis por subgrupos del ensayo de Kyriakopoulos *et al* (65) es mostrado en la **Tabla 5**.

El trabajo de Clarke *et al* se publicó de forma online el 27 de septiembre de 2019 (66). Se trata de un trabajo en el que se actualizaron los resultados descritos en el estudio STAMPEDE, que mostró resultado favorable en SG para docetaxel + ADT (73). Primero, los criterios para la interpretación de análisis por subgrupos de Sun *et al* fueron aplicados (4). Se estimó una $p(i) = 0,827$ para SG a partir de los datos disponibles de Clarke *et al*. Por ello, no se asumió heterogeneidad entre los subgrupos de elevado y bajo volumen tumoral. Estos subgrupos según la extensión de la enfermedad se encontraban pre-especificados en el estudio de actualización de los datos. La plausibilidad que podría respaldar una hipotética diferencia entre los subgrupos de alto y bajo volumen metastásico se ha mencionado en los estudios evaluados anteriormente: observaciones aparentes de estudios previos, posibles diferencias biológicas entre tumores más y menos extendidos o impacto del uso de docetaxel en líneas de tratamiento posteriores (74,75). La ausencia de interacción o heterogeneidad entre los subgrupos para SG tuvo consistencia interna con los resultados obtenidos en otras variables analizadas en Clarke *et al*, como SLP [$p(i) = 0.855$] y el tiempo hasta progresión de metástasis, nuevas metástasis o muerte [$p(i) = 0.960$]. Sin embargo, se observó inconsistencia externa de los resultados descritos en Clarke *et al* con los observados en Kyriakopoulos *et al* (con heterogeneidad entre subgrupos) y Gravis *et al* (con resultado global negativo) (65,78). Por otra parte, no se observó heterogeneidad entre los subgrupos según volumen tumoral en ensayos sobre el uso de apalutamida o abiraterona en cáncer de próstata (79,80).

Por último, se utilizó nuestra herramienta validada de aplicabilidad clínica para el análisis por subgrupos (55). Como en casos anteriores, se contestó negativamente a la cuestión preliminar sobre la existencia de diferencia de efecto entre los subgrupos según el volumen de metástasis [$p(i) > 0.1$], por lo que no se continuó empleando la herramienta y se descartó la aplicabilidad del análisis por subgrupos. Un resumen de la interpretación del análisis por subgrupos del estudio de Clarke *et al* puede consultarse en la **Tabla 5** (66).

Tabla 5. Resumen de la interpretación del análisis por subgrupos de los estudios incluidos

Metodología	Criterios		Sweeney <i>et al</i> (estudio CHARTED)	Gravis <i>et al</i> (análisis post hoc de estudio GETUG-AFU 15)	Kyriakopoulos <i>et al</i> (actualización de datos de estudio CHARTED)	Clarke <i>et al</i> (actualización de datos de estudio STAMPEDE)
Sun <i>et al</i>	Interacción estadística		No [p(i) = 1]	No [p(i) = 0.4]	Sí [p(i) = 0.03]	No [p(i) = 0.827]
	Pre-especificación de subgrupos analizados		No, pero incluido en estratificación	No	Sí, e incluido en estratificación	Sí
	Plausibilidad biológica		Sí	Sí	Sí	Sí
	Consistencia de los resultados		Sí (consistencia interna)	No	No	No
Herramienta validada (Gil-Sierra <i>et al</i>)	Cuestiones preliminares	Nivel más alto de evidencia del estudio que contiene el análisis por subgrupos	Sí	Sí	Sí	Sí
		Relevancia clínica clara de la variable seleccionada o variable primaria subrogada del estudio	Sí	Sí	Sí	Sí
		Existencia de diferencia de efecto entre los subgrupos para el factor considerado [p(i) < 0.1]	No	No	Sí	No
		Factor determinante del análisis por subgrupos previo a la intervención sanitaria	No aplicado	No aplicado	Sí	No aplicado
	Checklist Asociación estadística (puntuación) Plausibilidad biológica (puntuación) Consistencia (puntuación) Recomendación		No aplicado No aplicado No aplicado No aplicado	No aplicado No aplicado No aplicado No aplicado	Aplicado Probable (+3 puntos) Probable (+3 puntos) Nula (-3 puntos) Nula: El análisis por subgrupos no debe aplicarse	No aplicado No aplicado No aplicado No aplicado

3.2.2.3.3. Riesgo de sesgo

El porcentaje de pacientes con puntuación de escala Gleason ≥ 8 en los ECA fue el siguiente: 56.1% en la publicación de Gravis *et al*, 61.2% en el estudio CHAARTED y 67.5% en el ensayo de Clarke *et al* (65,66,78). Las medianas de los niveles de PSA (brazo con intervención – brazo control) fueron 26.7–25.8 ng/ml en el estudio de Gravis *et al*, 50.9–52.1 ng/ml en el estudio CHAARTED y 97–102.5 ng/ml en el estudio de Clarke *et al*. Los porcentajes de pacientes con estado funcional ECOG 1–2 fueron: 2.3% en la publicación de Gravis *et al*, 30.5% en el ensayo CHAARTED y 27.2% en estudio de Clarke *et al*. Los ECA mostraron los siguientes porcentajes de pacientes con mHSPC que no habían sido tratados previamente: 70.6% en el estudio de Gravis *et al*, 72.8% en el estudio CHAARTED y 95.4% en el estudio de Clarke *et al*. El subgrupo de pacientes con baja carga metastásica incluyó a 202 pacientes en la publicación de Gravis *et al*, 277 en el ensayo CHAARTED y 362 en el estudio por Clarke *et al*. El subgrupo de pacientes con alta carga metastásica englobó a 183 pacientes en el estudio de Gravis *et al*, 513 en el ensayo CHAARTED y 468 en el estudio de Clarke *et al*.

3.2.2.4. Discusión

En este trabajo no se han observado diferencias consistentes entre los subgrupos de alto y bajo volumen metastásico en los pacientes con mHSPC tratados con docetaxel. A día de hoy, no existe suficiente evidencia científica para descartar el beneficio en SG aportado por docetaxel en pacientes con mHSPC y carga tumoral baja. En algunos estudios de este contexto clínico se ha incurrido en errores de interpretación del análisis por subgrupos, como asumir diferencias entre subgrupos al comparar valores de medianas, o interpretando gráficas sin considerar siquiera $p(i)$ (64,74,75). En este caso, se podría estar privando a pacientes con mHSPC de recibir docetaxel, el cual es un fármaco que aporta un importante beneficio en SG. Estos hallazgos demuestran que los análisis por subgrupos deben interpretarse con cautela, aunque la hipótesis inicial de una posible diferencia de efecto pudiera parecer razonable.

Como ya se ha comentado previamente, la aplicación del análisis por subgrupos presenta una serie de limitaciones. Por ello, debe ser llevada a cabo prudentemente y con criterios metodológicos estrictos. El incremento de los errores α y β inherente al análisis por subgrupos debe ser considerado en la toma de decisiones clínicas (5). Por ello, es imprescindible disponer de un considerable reclutamiento de pacientes. La falta de potencia estadística pudo comprobarse en el estudio de Gravis *et al* (78), donde los autores asumieron que el tamaño muestral no fue lo suficientemente grande para poder determinar posibles diferencias entre los subgrupos según la carga de la enfermedad. De hecho, se detalló que se registró un número insuficiente de muertes en el subgrupo de bajo volumen tumoral. También hubo otra limitación importante: el carácter retrospectivo *post hoc* de la publicación. Este tipo de diseño sería susceptible de interpretar de manera artificiosa determinados resultados de algunas intervenciones sanitarias. Por otra parte, los pacientes incluidos en Gravis *et al* presentaron mejor pronóstico de la enfermedad (menor puntuación en la escala Gleason, menor PSA, menor estado general, etc.) que la población incluida en los estudios CHARTED y de Clarke *et al* (64,66). Esto explicaría las diferencias de medianas de OS existentes entre estos estudios.

Clarke *et al* y Kyriakopoulos *et al* no presentaron la limitación de un tamaño muestral escaso (65,66). Clarke *et al* no encontró diferencia estadísticamente significativa entre los subgrupos de alto y bajo volumen metastásico en mHSPC, mientras que Kyriakopoulos *et al* sí lo hizo. Clarke *et al* fue el estudio con mayor número de pacientes e incluyó un 95% de pacientes con metástasis *de novo*, con un número de pacientes considerable en el subgrupo de baja carga tumoral. Sin embargo, aproximadamente el 27-30% de los pacientes incluidos en los ensayos CHARTED y GETUG-15 presentaron enfermedad metastásica después de tratamiento previo radical. En este caso, los subgrupos de baja carga tumoral incluyeron un número limitado de pacientes con mHSPC *de novo* (64,77). Estas diferencias influyen en el pronóstico y evolución de los pacientes (81), y deben ser tenidas en cuenta en la toma de decisiones clínicas. La aplicación del análisis por subgrupos debe considerar las características basales de los

ECA. Por tanto, Clarke *et al* se presenta como el estudio más robusto y contradice los resultados obtenidos en otros estudios con mayores limitaciones

Considerando los datos de subgrupos y el beneficio en SG obtenido por docetaxel, parece razonable afirmar que este fármaco ha demostrado ser un tratamiento efectivo en la primera línea de tratamiento para pacientes con mHSPC, independientemente de la carga tumoral. Esta alternativa terapéutica no debería ser excluida del arsenal terapéutico usado en el subgrupo de pacientes con bajo volumen metastásico en mHSPC. Asimismo, existe bibliografía que respalda que no tiene un impacto negativo a largo plazo en la calidad de vida de los pacientes y que es una alternativa válida a abiraterona, apalutamida o enzalutamida (82). De esta manera, se mantendría un mayor número de alternativas terapéuticas posibles en el contexto clínico de mHSPC, favoreciendo la competencia de precios y la eficiencia en los sistemas de salud.

3.2.3. Uso temprano de remdesivir en pacientes con COVID-19.

Gil-Sierra MD, Briceño-Casado MP, Alegre-Del Rey EJ, Sánchez-Hidalgo M. Efficacy of early use of remdesivir: a systematic review of subgroup analysis. *Rev Esp Quimioter.* 2022;35(3):249-259. Doi: 10.37201/req/154.2021.

3.2.3.1. Introducción

La COVID-19 ha afectado a 200 millones de personas en todo el mundo, llegando a producir más de 4 millones de fallecidos (83). Debido a la situación de emergencia, se usaron de forma experimental diferentes fármacos para el tratamiento de la infección y muchos estudios fueron desarrollados (84,85). Sólo las vacunas y corticoides han demostrado un claro beneficio relevante (86–89).

Remdesivir es un profármaco de un análogo de adenosina que ha demostrado actividad *in vitro* contra el SARS-CoV-2. Los resultados clínicos del antiviral han sido evaluados en algunos ECA. El estudio de Wang *et al* comparó remdesivir frente a placebo, incluyendo a pacientes de Wuhan (31). En este ECA no se observó diferencia estadísticamente significativa en ninguna de las variables analizadas, entre las que se encuentran la mortalidad a los 28 días y el tiempo hasta mejoría clínica. A pesar de ello, se sugirió un posible beneficio en el tiempo hasta mejoría clínica en aquellos pacientes tratados de forma temprana de acuerdo a los resultados de un análisis por subgrupos (**Tabla 6**).

Tabla 6. Resultados iniciales del uso temprano de remdesivir. Tomada de Wang et al. (2020) (31).

	Remdesivir group (n=158)	Placebo group (n=78)	Difference*
Time to clinical improvement	21.0 (13.0 to 28.0)	23.0 (15.0 to 28.0)	1.23 (0.87 to 1.75)†
Day 28 mortality	22 (14%)	10 (13%)	1.1% (-8.1 to 10.3)
Early (≤10 days of symptom onset)	8/71 (11%)	7/47 (15%)	-3.6% (-16.2 to 8.9)
Late (>10 days of symptom onset)	12/84 (14%)	3/31 (10%)	4.6% (-8.2 to 17.4)
Clinical improvement rates			
Day 7	4 (3%)	2 (3%)	0.0% (-4.3 to 4.2)
Day 14	42 (27%)	18 (23%)	3.5% (-8.1 to 15.1)
Day 28	103 (65%)	45 (58%)	7.5% (-5.7 to 20.7)

La aplicación del análisis por subgrupos presenta una serie de limitaciones, por lo que debe ser llevada a cabo prudentemente y con criterios metodológicos sólidos (4). La comercialización de remdesivir y la crisis sanitaria por COVID-19 ha desencadenado un aluvión de hipótesis y afirmaciones con resultados inmaduros de fármacos, en algunos casos sobre análisis por subgrupos. Algunas de estas hipótesis abordan el momento de uso del antiviral, asociando el uso temprano de remdesivir con un mayor beneficio para los pacientes (31,90). Sin embargo, ¿ha demostrado el uso temprano de este fármaco un beneficio estadísticamente significativo? ¿En qué variables? ¿Cuál es el corte real para diferenciar el uso temprano y uso tardío de remdesivir? Las respuestas a estas preguntas son controvertidas y un excesivo –e inadecuado- uso de remdesivir podría afectar considerablemente a la eficiencia del sistema sanitario público. Asimismo, un análisis detallado de este tema podría minimizar futuros errores de interpretación en

situaciones de emergencia similares. El objetivo de este trabajo es la interpretación de los análisis por subgrupos según el momento de uso de remdesivir con respecto al inicio de los síntomas en pacientes con COVID-19, usando criterios sistemáticos y metodologías establecidos.

3.2.3.2. Métodos

3.2.3.2.1. Búsqueda bibliográfica

Se utilizó la estrategia PICOS (53) para formular la pregunta de investigación: población, intervención, comparador, resultados y diseño de estudios. Se incluyeron pacientes con COVID-19 como población. El tratamiento con remdesivir o combinaciones de fármacos con este antiviral fueron seleccionados como intervención. Se incluyó cualquier comparador. El resultado objetivo fueron análisis por subgrupos según el inicio de los síntomas para cualquier variable. ECA fue el diseño de estudio seleccionado.

La guía PRISMA fue aplicada en la búsqueda sistemática (54). La revisión fue llevada a cabo en la base de datos Pubmed® a fecha de 20 de agosto de 2021. La estrategia de búsqueda se desarrolló usando la herramienta 'Clinical Queries/Narrow': (*remdesivir AND covid*) AND (*Therapy/Narrow[filter]*). Posteriormente, se llevó a cabo un rastreo de citas.

3.2.3.2.2. Cribado y selección de estudios

La búsqueda fue desarrollada por dos investigadores de forma independiente. Las incongruencias en los resultados de la búsqueda fueron resueltas por discusión. Los títulos y resúmenes de los registros de la búsqueda fueron revisados para identificar y descartar estudios que no cumplieran los criterios de inclusión. El texto completo de los resultados de la búsqueda fue examinado en el proceso de elegibilidad. ECA sobre COVID-19 con análisis por subgrupos según el inicio de los síntomas fueron

seleccionados. El uso temprano de remdesivir fue definido como el inicio del tratamiento antes de un día concreto, teniendo en cuenta el inicio de los síntomas. Este corte para identificar el uso temprano del antiviral fue detallado en cada estudio. El uso tardío de remdesivir fue aquel que ocurre fuera del período que comprende el uso temprano. Los registros obtenidos en la revisión en un idioma distinto del español o el inglés fueron excluidos.

3.2.3.2.3. Extracción de datos

Los ECA incluidos en este trabajo se evaluaron en orden según la fecha de publicación. Se recogieron los siguientes datos: autores, fecha de publicación, población, terapias usadas como intervención y comparador, tamaños muestrales, variables, número de días que divide a la población global del estudio en subgrupos y eficacia de los tratamientos.

3.2.3.2.4. Análisis de datos

Los criterios de evaluación del análisis por subgrupos fueron los descritos en Sun *et al* (4): interacción estadística, pre-especificación de subgrupos, plausibilidad biológica que respalde al efecto observado y consistencia de los resultados de un ECA con respecto a otros estudios o datos descritos dentro del mismo ECA (consistencia interna). Considerando las limitaciones del análisis por subgrupos, se establecen como límites aceptables de heterogeneidad a valores de $p(i) < 0.1$ (2). Se llevó a cabo el cálculo de $p(i)$ para aquellos estudios sin este dato mediante los valores de riesgo relativo, odds ratio, hazard ratio e ICs proporcionados (12,68–72). Si no se disponía de suficientes datos y siempre que fuera posible, $p(i)$ fue valorado usando las gráficas y tablas.

Posteriormente, se usó nuestra herramienta validada para valorar la aplicabilidad del análisis por subgrupos mediante una metodología ordenada y sistemática (55). Esta herramienta se estructura en dos partes: unas cuestiones preliminares que permiten descartar la valoración de los análisis por subgrupos sin condiciones mínimas relevantes, y un cuestionario. Estas cuestiones preliminares abordan el nivel de evidencia del

estudio que contiene el análisis por subgrupos, la relevancia de la variable seleccionada, la existencia de diferencia de efecto [$p(i) < 0.1$] entre los subgrupos para el factor considerado, y la secuencia temporal entre este factor y la intervención sanitaria. La respuesta negativa de alguna de las cuestiones preliminares conlleva al descarte de la realización de las siguientes preguntas, del cuestionario y la aplicabilidad del análisis por subgrupos. En el cuestionario se otorga una valoración y puntuación a cada uno de los criterios establecidos para la interpretación del análisis por subgrupos: asociación estadística, que incluye $p(i)$, pre-especificación de subgrupos, tamaño muestral, número de factores estudiados y resultado global del estudio; plausibilidad biológica y consistencia de los resultados. La puntuación es más alta cuanto mayor sea la fiabilidad del criterio: probable (+3 puntos), posible (+2), dudosa (0) y nula (-3). El sumatorio de las puntuaciones de estos criterios se relaciona con una recomendación para la aplicabilidad de los resultados en la práctica clínica. Puntuaciones más altas del cuestionario se corresponden con una mayor aplicabilidad de los resultados del análisis por subgrupos y, por tanto, con recomendaciones para que deban tenerse en cuenta en la toma de decisiones clínicas. Los sumatorios de las puntuaciones se asocian a las siguientes recomendaciones: 9-7 puntos (recomendación probable) se corresponde con aplicabilidad del análisis por subgrupos; 5-6 puntos (recomendación posible) con aplicabilidad prudente ante escasa tolerancia, dificultad de uso o coste de otras intervenciones; 3-4 puntos (recomendación dudosa) se relaciona con evitar aplicabilidad salvo excepciones; y <3 puntos (recomendación nula) se relaciona con la no aplicación del análisis por subgrupos.

3.2.3.2.5. Evaluación del riesgo de sesgo

Los resultados de los ECA pueden verse afectados por el pronóstico de los pacientes reclutados. Se revisaron la gravedad de la enfermedad, existencia de comorbilidades y puntuación basal en las escalas ordinales sobre el uso de oxígeno de los pacientes incluidos en los estudios. También se analizaron los tamaños muestrales de los subgrupos descritos en los ECA.

3.2.3.3. Resultados

3.2.3.3.1. Búsqueda de la literatura y características de los estudios

Se obtuvieron cincuenta y tres resultados en la búsqueda sistemática usando la base de datos Pubmed®. En el rastreo de citas se encontró un ECA sin análisis de subgrupos que comparó los regímenes de 5 y 10 días de remdesivir (91). Se excluyeron los registros según la siguiente distribución de los criterios de exclusión: 36 estudios presentaron un diseño diferente del ECA, 9 evaluaron un fármaco diferente a remdesivir, 3 no contenían análisis por subgrupos y un estudio no abordó resultados de eficacia. Del total inicial de 54 registros, se incluyeron 5 ECA. La **Figura 8** ilustra la revisión de la literatura mediante el protocolo PRISMA y la **Tabla 7** muestra los datos de los estudios incluidos.

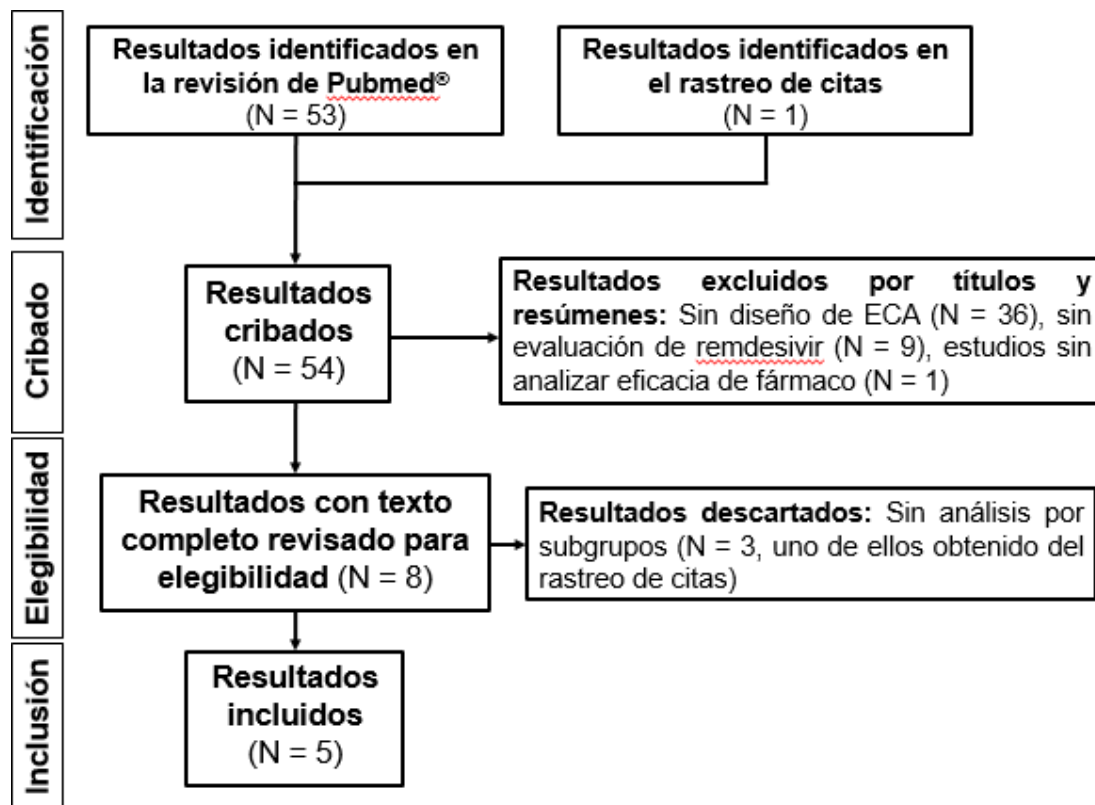


Figura 8. Revisión sistemática de la literatura

Tabla 7. Datos de los ensayos clínicos incluidos

Autores	Fecha de publicación online	Población	Intervención	Comparador	Tamaño muestral del ensayo	Variables evaluadas con análisis por subgrupos	Subgrupos según el tiempo desde inicio de síntomas hasta el tratamiento	Tamaño muestral del subgrupo con uso temprano de remdesivir†	Tamaño muestral del subgrupo con uso tardío de remdesivir†	Eficacia para variables en la población global (95% IC)	Eficacia en subgrupo con uso temprano de remdesivir (95% IC)	Eficacia en subgrupo con uso tardío de remdesivir (95% IC)
Wang <i>et al.</i> (31)	29 de abril de 2020	Pacientes con COVID-19 grave	Remdesivir 200 mg el día 1 y remdesivir 100 mg en los días 2-10	Placebo	237	Tiempo hasta mejoría clínica	≤10 días y >10 días	118	115	Hazard Ratio: 1.23 (95% IC 0.87 a 1.75)	Hazard Ratio: 1.52 (95% IC 0.95 a 2.43)	Hazard Ratio: 1.07 (95% IC 0.63 a 1.83)
						Mortalidad en el día 28	≤10 días y >10 días	118	115	Diferencias de tasas: 1.1% (95% IC -8.1 a 10.3)	Diferencias de tasas: -3.6% (95% IC -16.2 a 8.9)	Diferencias de tasas: 4.6% (95% IC -8.2 a 17.4)
						Carga viral de RNA en el tracto respiratorio alto	≤10 días y >10 días	64	60	Sin datos de carga viral -Log10 copias/mL (solo gráficas disponibles)	Sin datos de carga viral -Log10 copias/mL (solo gráficas disponibles)	Sin datos de carga viral -Log10 copias/mL (solo gráficas disponibles)
Spinner <i>et al.</i> (92)	21 de agosto de 2020	Pacientes con COVID-19 moderado	Remdesivir 200 mg el día 1 y remdesivir 100 mg en los días 2-10 (esquemas de 5 y 10 días)	Cuidados estándares	596	Estado clínico en el día 11	Esquema de tratamiento de 5 días: <9 días y ≥9 días	Sin datos	Sin datos	Sin datos de diferencia en proporciones (solo gráficas disponibles)	Sin datos de diferencia en proporciones (solo gráficas disponibles)	Sin datos de diferencia en proporciones (solo gráficas disponibles)
						Esquema de tratamiento de 10 días: <9 días y ≥9 días	Sin datos	Sin datos	Sin datos de diferencia en proporciones (solo gráficas disponibles)	Sin datos de diferencia en proporciones (solo gráficas disponibles)	Sin datos de diferencia en proporciones (solo gráficas disponibles)	
Beigel <i>et al.</i> (Grupo del estudio ACTT-1, informe final) (93)	9 de octubre de 2020	Pacientes con COVID-19 grave	Remdesivir 200 mg el día 1 y remdesivir 100 mg en los días 2-10	Placebo	1062	Tiempo hasta mejoría clínica	≤10 días y >10 días	676	383	Relación de tasas: 1.29 (95% IC 1.12 a 1.49)	Relación de tasas: 1.37 (95% IC 1.14 a 1.64)	Relación de tasas: 1.20 (95% CI 0.94 a 1.52)
							≤9 días y >9 días	582	477	Relación de tasas: 1.29 (95% IC 1.12 a 1.49)	Relación de tasas: 1.32 (95% IC 1.09 a 1.61)	Relación de tasas: 1.29 (95% IC 1.04 a 1.59)
							1 ^{er} : <7 días‡	282	777	Relación de tasas: 1.29 (95% IC 1.12 a 1.49)	Relación de tasas: 1.92 (95% IC 1.41 a 2.60)	-
							2° cuartil: 7 a ≤9 días‡	300	-	Relación de tasas: 1.29	Relación de tasas: 0.99	-

										(95% IC 1.12 a 1.49)	(95% IC 0.76 a 1.28)		
							3 ^{er} cuartil: 10 a ≤ 12 días†	221	-	Relación de tasas: 1.29 (95% IC 1.12 a 1.49)	Relación de tasas: 1.45 (95% IC 1.07 a 1.98)	-	
							4 ^o cuartil: ≥13 días‡	803	256	Relación de tasas: 1.29 (95% IC 1.12 a 1.49)	Relación de tasas: 1.15 (95% IC 0.86 a 1.54)	-	
						Mejor (menor) escala de estado clínico en el día 15	≤10 días y >10 días	676	383	Odds ratio: 1.6 (95% IC 1.3 a 1.9)	Odds ratio: 1.7 (95% IC 1.3 a 2.2)	Odds ratio: 1.3 (95% IC 0.9 a 1.9)	
							≤9 días y >9 días	582	477	Odds ratio: 1.6 (95% IC 1.3 a 1.9)	Odds ratio: 1.7 (95% IC 1.2 a 2.2)	Odds ratio: 1.4 (95% IC 1.0 a 1.9)	
							1 ^o cuartil: <7 días‡	282	777	Odds ratio: 1.6 (95% IC 1.3 a 1.9)	Odds ratio: 2.7 (95% IC 1.8 a 4.2)	-	
							2 ^o cuartil: 7 a ≤ 9 días‡	300	-	Odds ratio: 1.6 (95% IC 1.3 a 1.9)	Odds ratio: 1.1 (95% IC 0.8 a 1.7)	-	
							3 ^o cuartil: 10 a ≤ 12 días‡	221	-	Odds ratio: 1.6 (95% IC 1.3 a 1.9)	Odds ratio: 1.4 (95% IC 0.9 a 2.3)	-	
							4 ^o cuartil: ≥13 días‡	803	256	Odds ratio: 1.6 (95% IC 1.3 a 1.9)	Odds ratio: 1.3 (95% IC 0.9 a 2.1)	-	
Kalil <i>et al.</i> (94)	5 de enero de 2021	Pacientes con COVID-19 moderado y grave	Remdesivir 200 mg el días 1 y remdesivir 100 mg en los días 2-10 + baricitinib 4 mg durante 14 days	Placebo + Remdesivir 200 mg el día 1 y remdesivir 100 mg en los días 2-10	1033		Tiempo hasta mejoría clínica	≤10 días y >10 días	764	253	Relación de tasas: 1.16 (95% IC 1.01 a 1.32)	Relación de tasas: 1.13 (95% IC 0.97 a 1.32)	Relación de tasas: 1.27 (95% IC 0.97 a 1.67)
Barratt-Due <i>et al.</i> (95)	13 de julio de 2021	Pacientes con COVID-19 moderado y grave	Remdesivir 200 mg el días 1 y remdesivir 100 mg hasta el día 9	Cuidados estándares	185		Aclaramiento viral orofaríngeo	<7 días y ≥7 días	Sin datos	Sin datos	Diferencia tasa disminución diaria: 0.113 (95% IC -0.001 a 0.227)	Diferencia tasa disminución diaria: 0.19 (95% IC 0.03 a 0.36)	Diferencia tasa disminución diaria: 0.02 (95% IC -0.15 a 0.19)

†El uso temprano de remdesivir fue relacionado con subgrupo de pacientes que recibieron el antiviral en el menor número de días entre el inicio de los síntomas y comienzo del tratamiento. El uso tardío de remdesivir es el subgrupo de pacientes con mayor número de días en ese período de tiempo. ‡Cuartiles sobre duración de síntomas previo al reclutamiento de pacientes en estudio del Grupo ACTT-1 solo presentaron un dato de eficacia. El subgrupo complementario en 1^{er} y 4^o cuartil fue estimado a partir de datos del ensayo. % IC= porcentaje de intervalo de confianza.

3.2.3.3.2. Análisis por subgrupos

Wang *et al* publicaron un análisis por subgrupos sobre el efecto del uso temprano y tardío de remdesivir (con un punto de corte a los 10 días desde el inicio de los síntomas) para las siguientes variables: tiempo hasta mejoría clínica, mortalidad en el día 28 y carga viral de ARN en las vías respiratorias superiores (31). Primero, se aplicaron los criterios para la interpretación de análisis por subgrupos descritos en Sun *et al* (4). El ensayo no aportó el valor de $p(i)$ para ninguna de las variables citadas. Se calculó y no se encontró heterogeneidad para el tiempo hasta la mejoría clínica entre los subgrupos de uso temprano y tardío de remdesivir [$p(i) = 0,33$] ni para la mortalidad en día 28 [$p(i) = 0,37$]. No obstante, Wang *et al* no proporcionó datos suficientes para calcular $p(i)$ entre estos subgrupos para la carga viral de ácido ribonucleico (ARN) en las vías respiratorias superiores. Los autores de este trabajo sólo comentaron que no se obtuvieron diferencias estadísticamente significativas entre los brazos control e intervención para la carga viral de ARN en las vías respiratorias superiores ni en el uso temprano ni tardío de remdesivir. No se mencionó el valor de $p(i)$ entre estos subgrupos. Asimismo, el protocolo del estudio no pre-especificó el análisis por subgrupos según el inicio de los síntomas para ninguna variable. Por otra parte, la plausibilidad biológica que justificaría el uso temprano de remdesivir en pacientes con COVID-19 parece razonable. Existe una hipótesis basada en la experiencia previa con virus como influenza u otros patógenos que argumenta que el uso temprano de remdesivir podría ejercer un mayor efecto antes de que el virus se expanda demasiado (90,96). Wang *et al* fue el primer ECA sobre el uso de remdesivir en COVID-19 y la consistencia con ensayos previos no pudo evaluarse. Sin embargo, la ausencia de heterogeneidad entre los subgrupos durante el tiempo hasta la mejoría clínica y la mortalidad en el día 28 mostró consistencia interna de los resultados del análisis de subgrupos (31).

A continuación, se usó nuestra herramienta validada sobre la aplicabilidad de análisis de subgrupos (55). Algunas de las preguntas preliminares fueron respondidas negativamente para las tres variables evaluadas, por lo que no se siguió con el cuestionario. El análisis por subgrupos según el momento de administración del fármaco para el tiempo hasta la mejoría clínica y la mortalidad en el día 28 no presentó

diferencias de efecto entre subgrupos [$p(i) > 0.1$]. Tampoco se proporcionaron datos suficientes para la estimación de la heterogeneidad entre subgrupos para la carga viral de ARN en las vías respiratorias superiores. De todas maneras, esta variable es una variable subrogada de escasa relevancia clínica. La **Tabla 8** muestra un resumen de la interpretación del análisis por subgrupos descrito en Wang *et al* (31).

Spinner *et al* llevaron a cabo un ECA con análisis de subgrupos que incluía el uso temprano y tardío de remdesivir (con un corte a los 9 días) para las siguientes variables: estado clínico en el día 11, tanto para el esquema de remdesivir de 5 días como para el de 10 días (92). Nuevamente, se aplicaron los criterios para la interpretación de análisis de subgrupos de Sun *et al* (4). El valor de $p(i)$ entre los subgrupos valorados no se describe para el estado clínico en el día 11 y no hay datos exactos reportados de los subgrupos. Sin embargo, el *forest plot* del análisis de subgrupos mostró una superposición casi total de los intervalos de confianza (IC) del uso temprano y tardío de remdesivir para el esquema de 5 días [$p(i) > 0,2$], mientras que la superposición fue menor que el 50% de los IC para el esquema de 10 días. La diferencia en proporciones y los IC del 95% se extrajeron de las representaciones gráficas de *forest plot* para el esquema de remdesivir de 10 días. La diferencia en proporciones para el uso temprano de remdesivir (<9 días) fue 12,6 (95% IC, -0,6 a 5,3), y para el uso tardío de remdesivir (≥ 9 días) fue -4 (95% IC, -18,6 a 10). Con estos datos se obtuvo el valor de $p(i)$ entre estos subgrupos, estimado en 0,0393. El análisis por subgrupos sobre el momento de uso de remdesivir con respecto al inicio de los síntomas no se pre-especificó previamente (análisis *post hoc*). Como ya hemos comentado antes, la plausibilidad biológica podría apoyar un hipotético mejor resultado del uso temprano de remdesivir, de acuerdo a experiencias previas con virus y otros microorganismos (90,96). Se observó inconsistencia entre los resultados de los subgrupos de los esquemas de 5 y 10 días de tratamiento dentro del mismo ensayo (92). Mientras que un esquema de remdesivir demostró diferencia estadísticamente significativa entre los subgrupos de uso temprano y tardío de remdesivir, el otro esquema no lo hizo. Este estudio tiene un corte diferente del momento de uso de remdesivir con respecto a Wang *et al*, el cual diferencia el uso temprano y tardío según se administre el fármaco antes o después de los 10 días,

respectivamente (4). Teniendo esto en cuenta, no se pueden realizar comparaciones entre subgrupos con diferentes cortes en los períodos de tiempo.

Posteriormente, usamos la herramienta que hemos desarrollado sobre la aplicabilidad del análisis de subgrupos (55). Para el esquema de 5 días de remdesivir, la pregunta preliminar sobre la existencia de diferencias de efecto entre los subgrupos según el momento de uso de remdesivir fue respondida negativamente para la variable del estado clínico en el día 11 [$p(i) > 0.2$]. Para el esquema de 10 días de remdesivir, todas las preguntas preliminares fueron respondidas positivamente. Al aplicar el cuestionario, la asociación estadística presentó una evaluación "nula" (-3 puntos) debido a la falta de pre-especificación del factor (análisis *post hoc*). La plausibilidad biológica obtuvo una evaluación "probable" (+3 puntos) y la consistencia fue "nula" (-3 puntos). La suma de estas puntuaciones (-3 puntos) se asoció con una "nula" recomendación de aplicabilidad en la práctica clínica de los resultados de subgrupos. Por lo tanto, los resultados del subgrupo de uso temprano (<9 días) del régimen de 10 días de remdesivir no debe considerarse para la toma de decisiones clínicas en COVID-19 moderado. Un resumen de la interpretación del análisis por subgrupos descrito en Spinner *et al* puede consultarse en la **Tabla 8**.

Beigel *et al* (informe final del Grupo del estudio ACTT-1) publicó un análisis por subgrupos según la duración de los síntomas antes del reclutamiento para dos variables: tiempo hasta mejoría clínica y mejor escala de estado clínico en el día 15 (93). El análisis por subgrupos evaluó estas variables de eficacia a través de los siguientes cortes: cuartiles de días (<7 días, de 7 a ≤ 9 días, de 10 a ≤ 12 días, >13 días); ≤ 9 días *versus* >9, y ≤ 10 días *versus* >10. Primero, los criterios de Sun *et al* se utilizaron para valorar el análisis por subgrupos (4). No se reportaron valores de $p(i)$ para las variables de eficacia en los diferentes cortes. Se realizaron las estimaciones, y se encontró heterogeneidad [$p(i)=0,009$] en el primer cuartil de días (subgrupo con <7 días de síntomas *versus* ≥ 7 días) para el tiempo hasta mejoría clínica. $P(i) < 0,05$ también fue estimado en el primer cuartil para una mejor escala del estado clínico en el día 15. No obstante, se cuantificó

un valor de $p(i) > 0,2$ para cualquiera de los otros cortes en ambas variables. Los análisis por subgrupos sobre la duración de los síntomas fueron pre-especificados en el protocolo del ensayo. Tal y como hemos comentado previamente, podría existir una plausibilidad biológica razonable que apoyara al uso temprano de remdesivir según la experiencia con otros microorganismos (90,96). Se encontró consistencia interna (diferencias estadísticamente significativas) entre los resultados de los subgrupos de uso temprano y tardío para el tiempo hasta la mejoría clínica y mejor escala de estado clínico en el día 15. Ambas variables presentaron mejoría estadísticamente significativa [$p(i) < 0,05$] para los pacientes con tratamiento temprano de remdesivir en < 7 días desde el inicio de los síntomas. Por otra parte, los resultados con cortes a los 9 y 10 días se compararon con datos publicados anteriormente (31,92). Ninguno de estos estudios mostró un beneficio consistente para el uso temprano de remdesivir con cortes en los días 9 y 10. El resto de subgrupos (7 a ≤ 9 días, de 10 a ≤ 12 días y > 13 días) no fueron evaluados en estudios previos y, por lo tanto, no se pudo llevar a cabo una comparación con los datos de Beigel *et al* (93).

A continuación, se utilizó nuestra herramienta sobre aplicabilidad del análisis de subgrupos (55). Las preguntas preliminares tuvieron respuestas positivas en el primer cuartil (< 7 días de síntomas *versus* ≥ 7 días) para el tiempo hasta la mejoría clínica y mejor escala de estado clínico en el día 15. Al aplicar el cuestionario, la asociación estadística mostró una valoración "probable" (+3 puntos). La plausibilidad biológica también presentó una evaluación "probable" (+3 puntos). La consistencia obtuvo una valoración "dudosa" (0 puntos). La suma de puntuaciones (+6 puntos) se asoció con una recomendación "posible" de aplicabilidad en la práctica clínica de los resultados de subgrupos. Por lo tanto, un mayor beneficio del uso temprano de remdesivir en pacientes con < 7 días de síntomas para el tiempo hasta mejoría clínica y mejor escala de estado clínico en el día 15 puede ser aplicado con precaución en la práctica clínica de COVID-19 grave. Las cuestiones preliminares sobre la existencia de diferencias de efecto entre subgrupos fueron respondidas negativamente para el resto de los días de corte en las variables seleccionadas. Un resumen de la interpretación del análisis por subgrupos

sobre el corte de 7 días de síntomas del informe final del Grupo del estudio ACTT-1 es mostrada en la **Tabla 8**.

Kalil *et al* compararon la combinación de baricitinib con remdesivir *versus* remdesivir (94). Este ECA incluyó un análisis por subgrupos según el inicio de los síntomas (con un corte durante 10 días) para la variable del tiempo hasta mejoría clínica. De nuevo, se aplicaron los criterios de análisis por subgrupos de Sun *et al* (4). No se presentó ninguna estimación de $p(i)$ de los subgrupos para la variable evaluada. Se calculó el valor de $p(i)$, no encontrándose heterogeneidad entre los subgrupos de uso temprano y tardío del tratamiento [$p(i) = 0,46$]. El análisis por subgrupos según la duración de los síntomas fue pre-especificado en el protocolo del estudio. La plausibilidad biológica que justificaría un uso temprano de remdesivir pudiera parecer razonable, de acuerdo a experiencia previa con virus y otros microorganismos (90,96). Estudios previos sobre el uso de remdesivir en COVID-19 que contenían análisis por subgrupos según el inicio de los síntomas (con punto de corte a los 10 días) no mostraron heterogeneidad del efecto del antiviral (31,93). Estos datos podrían ser consistentes con los resultados de los subgrupos reportados en Kalil *et al* (94).

Posteriormente, se usó nuestra herramienta validada de aplicabilidad clínica para subgrupos (55). La pregunta preliminar sobre la existencia de diferencias de efecto entre subgrupos de uso temprano y tardío de remdesivir recibió una respuesta negativa [$p(i) > 0.1$]. Por lo tanto, no se aplicó el resto de la herramienta. La **Tabla 8** describe un resumen de la interpretación sobre el análisis por subgrupos detallado en Kalil *et al*.

Barratt-Due *et al* presentaron un análisis por subgrupos sobre el momento del uso de remdesivir (con corte a los 7 días desde inicio de los síntomas) para la variable de aclaramiento viral orofaríngeo (95). Se aplicaron los criterios de Sun *et al* para la interpretación del análisis por subgrupos (4). El ensayo no detalló el valor de $p(i)$ para la variable de aclaramiento viral orofaríngeo. Las estimaciones propias no mostraron heterogeneidad entre los subgrupos de uso temprano y tardío de remdesivir para la

variable seleccionada [$p(i) = 0,166$]. El análisis por subgrupos según el inicio de los síntomas fue pre-especificado en el protocolo del ensayo. La existencia de plausibilidad biológica para un hipotético beneficio del uso temprano de remdesivir para COVID-19 podría estar respaldada por la experiencia con otros patógenos infecciosos (90,96). La ausencia de heterogeneidad para la variable entre los subgrupos de uso temprano y tardío de remdesivir con el corte de 7 días para la variable de aclaramiento viral orofaríngeo de Barratt-Due *et al* mostró inconsistencia en comparación con las diferencias estadísticamente significativas observadas en las variables clínicas de Beigel *et al* (93,95).

A continuación, se empleó la herramienta que hemos desarrollado sobre el análisis por subgrupos (55). La pregunta preliminar sobre la relevancia de la variable evaluada en los subgrupos seleccionados tuvo una respuesta negativa, por lo que el resto del cuestionario no fue aplicado. Un resumen de la interpretación sobre el análisis por subgrupos reportado en Barratt-Due *et al* se detalla en la **Tabla 8**.

Tabla 8. Resumen de interpretaciones de análisis por subgrupos de los estudios incluidos.

Metodología	Criterios		Wang <i>et al.</i> (31)			Spinner <i>et al.</i> (92)		Beigel <i>et al.</i> (Grupo de estudio ACTT-1) (93)		Kalil <i>et al.</i> (94)	Barratt-Due <i>et al.</i> (95)	
			Variables			Variables		Variables con corte a 7 días de síntomas		Variables	Variables	
			Tiempo hasta mejoría clínica	Mortalidad en el día 28	Carga viral en tracto respiratorio superior	Estado clínico en el día 11		Tiempo hasta mejoría clínica	Mejor escala de estado clínico en el día 15	Tiempo hasta mejoría clínica	Aclaramiento viral orofaríngeo	
						Esquema 5 días de remdesivir	Esquema 10 días de remdesivir					
Sun <i>et al.</i> (4)	Interacción estadística		No	No	Datos insuficientes	No	Sí	Sí	Sí	No	No	
	Pre-especificación de subgrupos		No definido	No definido	No definido	No	No	Sí	Sí	Sí	Sí	
	Plausibilidad biológica		Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí	
	Consistencia de los resultados		Resultados con consistencia interna de no beneficio	Resultados con consistencia interna de no beneficio	Datos insuficientes	No	No	Sí (consistencia interna de mejor beneficio para uso temprano de remdesivir)	Sí (consistencia interna de mejor beneficio para uso temprano de remdesivir)	Sí (consistencia de ausencia de diferencias entre uso temprano y tardío de remdesivir)	No	
Herramienta validada (Gil-Sierra <i>et al.</i>) (55)	Cuestiones preliminares	Nivel más alto de evidencia del estudio con análisis por subgrupos	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí	
		Relevancia clínica clara de variable seleccionada o variable primaria subrogada	Sí	Sí	No	Sí	Sí	Sí	Sí	Sí	No	
		Diferencia de efecto entre subgrupos para factor considerado [$p(i) < 0.1$]	No	No	No aplicado	No	Sí	Sí	Sí	Sí	No	No aplicado
		Factor determinante del análisis por subgrupos previo a la intervención sanitaria	No aplicado	No aplicado	No aplicado	No aplicado	Sí	Sí	Sí	Sí	No aplicado	No aplicado
	Checklist Asociación estadística (puntuación) Plausibilidad biológica (puntuación) Consistencia (puntuación) Recomendación		No aplicado	No aplicado	No aplicado	No aplicado	Aplicado Nula (-3 puntos) Probable (+3 puntos) Nula (-3 puntos) Posible (6 puntos): "Análisis por subgrupos no deben ser considerados"	Aplicado Probable (+3 puntos) Probable (+3 puntos) Dudosa (0 puntos) Posible (6 puntos): "Análisis por subgrupos pueden ser aplicados a práctica clínica con cautela"	Aplicado Probable (+3 puntos) Probable (+3 puntos) Dudosa (0 puntos) Posible (6 puntos): "Análisis por subgrupos pueden ser aplicados a práctica clínica con cautela"	No aplicado	No aplicado	

3.2.3.3.3. Riesgo de sesgo

Tres ECA incluyeron poblaciones con COVID-19 moderada y severa, un ECA evaluó solo pacientes con COVID-19 moderada, y los demás reclutaron únicamente pacientes con enfermedad grave. El porcentaje de pacientes con comorbilidades en los ECA incluidos en este estudio oscila de 71% a 87,1%. La hipertensión fue la comorbilidad más frecuente en todos los ECA, con valores que van desde el 24,6% al 52% de los pacientes. La puntuación basal de los pacientes valorados con escalas ordinales sobre el uso de oxígeno presentó peores resultados en Beigel *et al* y Kalil *et al* (93,94), donde los porcentajes de pacientes con dispositivos de oxígeno de alto flujo o ventilación no invasiva fue del 17,6% al 21,8%, y los pacientes con ventilación mecánica invasiva u oxigenación por membrana extracorpórea varió desde el 10,5% y el 29,6%. Los ECA utilizaron escalas ordinales diferentes: un estudio aplicó una escala ordinal de seis categorías, otro evaluó pacientes con una escala de siete categorías, dos ECA usaron una escala de ocho categorías y uno aplicó una escala de diez categorías. Los datos sobre los factores relacionados con el pronóstico de los pacientes fueron detallados en la **Tabla 9**.

Wang *et al* analizó a más de 200 pacientes en el análisis por subgrupos según el momento de uso del fármaco antiviral para las variables de tiempo hasta mejoría clínica y mortalidad en el día 28; y menos de 130 pacientes para la carga viral de ARN en el tracto respiratorio superior (31). Más de 1.000 pacientes fueron incluidos en el análisis por subgrupos de Beigel *et al* y Kalil *et al* para el tiempo hasta la mejoría clínica y mejor escala de estado clínico en el día 15 (93,94). El resto de variables en los ECA presentaron datos insuficientes sobre los tamaños de la muestra de los subgrupos. La **Tabla 7** proporciona el número de pacientes involucrados en el análisis por subgrupos.

Tabla 9. Datos sobre los factores relacionados con el pronóstico de los pacientes en estudios.

Factores relacionados con el pronóstico de los pacientes	Wang <i>et al.</i> (31)	Spinner <i>et al.</i>† (92)	Beigel <i>et al.</i> (Grupo de estudio ACTT-1) (93)	Kalil <i>et al.</i> (94)	Barratt-Due <i>et al.</i> (95)
Gravedad de enfermedad (%) - Moderada - Severa	0% 100%	100% 0%	9,9% 90,1%	31,7% 68,3%	Moderada y severa Datos insuficientes Datos insuficientes
Comorbilidades (%) - Hipertension - Diabetes - Enfermedad cardiovascular - Obesidad	71% 46%-38% 25%-21% 9%-3%‡ Datos insuficientes	Datos insuficientes 44%-43%-41% 44%-37%-38% 58%-58%-54% Datos insuficientes	81,7%-82,2% 50,6%-50,9% 32%-31% 18%-16% 46%-45%	87,1%-81,7% 51%-52% 40%-36% 21%-21% 58%-53%	Datos insuficientes 36,6%-24,6% 22%-15,8% 14,6%-21,1% 28,9%-18,4%
Puntuación basal en escalas ordinales sobre uso de O ₂ (%) - Hospitalizado, requiriendo oxígeno suplementario - Hospitalizado, recibiendo ventilación no invasiva u oxígeno de alto flujo - Hospitalizado, recibiendo ventilación mecánica invasiva u oxigenación por membrana extracorpórea	82%-83% 18%-12% 0%-1%	12%-16%-19% 0% 0%	42,9%-39% 17,6%-18,8% 24,2%-29,6%	55,9%-53,3% 20%-21,8% 10,5%-11%	Datos insuficientes Datos insuficientes Datos insuficientes
Escala ordinal	Escala de 6 categorías	Escala de 7 categorías	Escala de 8 categorías	Escala de 8 categorías	Escala de 10 categorías

Los datos con un resultado detallan el porcentaje de población global del estudio. Los datos con dos resultados representan los porcentajes de los brazos intervención y control (% de pacientes en el brazo de intervención-% de los pacientes en el brazo de control). †Spinner *et al* incluyó tres brazos (92): esquema de 10 días de remdesivir, esquema de 5 días de remdesivir y brazo control (% de pacientes en brazo de esquema de 10 días de remdesivir -% de pacientes en brazo de esquema de 5 días de remdesivir-% de pacientes en brazo de control). ‡Sólo enfermedad coronaria.

3.2.3.4. Discusión

Los resultados en pacientes con COVID-19 severa sugirieron que el uso temprano de remdesivir proporciona beneficios en comparación con el uso tardío solo para las variables de tiempo hasta mejoría clínica y mejor escala de estado clínico en el día 15 (93). El efecto de remdesivir en estas variables clínicas ha sido estadísticamente superior en pacientes con <7 días de síntomas. Sin embargo, Barratt-Due *et al* no encontraron diferencias entre el uso temprano y tardío de remdesivir con corte a los 7 días de síntomas para el aclaramiento viral orofaríngeo (95). El aclaramiento viral orofaríngeo es una variable subrogada con escasa relevancia clínica. Se trata de una medida local y la COVID-19 severa es una enfermedad sistémica con afectación pulmonar. Esto podría ser una explicación de la inconsistencia entre los resultados de estas diferentes variables sobre el uso temprano y tardío de remdesivir con corte a los 7 días observado en Barratt-Due *et al* y Beigel *et al* (93,95). Mientras que es posible que no se encuentren diferencias en el aclaramiento de la carga viral en una parte del cuerpo como puede ser la cavidad orofaríngea, la evolución clínica de la enfermedad puede ser diferente al usar remdesivir de forma temprana por la presencia de una diseminación sistémica del virus. Por otra parte, no se encontraron hallazgos sobre un mayor efecto del uso temprano de remdesivir para otras variables y cortes de días con síntomas. De esta manera, la experiencia previa de un mayor beneficio del tratamiento temprano con neuraminidasa en pacientes con infección por influenza fue parcialmente confirmada en el uso temprano de remdesivir contra COVID-19 (97), ya que se observó beneficio sólo en variables con limitada relevancia clínica.

Los resultados de nuestro trabajo son complementarios a los resultados del ensayo SOLIDARITY de la OMS (98), que es el estudio más importante sobre tratamientos contra la COVID-19. Los autores del ensayo SOLIDARITY de la OMS concluyeron que remdesivir mostró escaso o ningún efecto en los pacientes hospitalizados con COVID-19, según datos de mortalidad, tiempo de hospitalización e inicio de la ventilación. Este estudio no fue incluido en nuestra revisión debido a la ausencia de análisis por subgrupos que evaluaran el efecto de remdesivir según el momento de uso teniendo en cuenta el inicio de los síntomas.

Para el punto de corte de 9 días de síntomas, nuestra revisión encontró una aparente heterogeneidad entre subgrupos en el esquema de 10 días de remdesivir para el estado clínico en el día 11 en COVID-19 moderado del ensayo de Spinner *et al* (92). Estas diferencias entre subgrupos con respecto al inicio de los síntomas fueron estimadas a partir de las representaciones gráficas, ya que los autores del estudio no aportaron datos exactos. Además, la naturaleza *post hoc* de este análisis por subgrupos no debe obviarse. Del mismo modo, Spinner *et al* y otros estudios similares no reportaron diferencias entre los subgrupos de uso temprano y tardío del régimen de remdesivir de 5 días para el corte de 9 días de síntomas (92,93). Las diferencias aparentes entre subgrupos en el esquema de 10 días de remdesivir pueden ser atribuidas a la multiplicidad de determinaciones y las limitaciones de los resultados descritos en subgrupos (5).

Los análisis de subgrupos deben ser considerados con precaución debido a sus limitaciones (4,5,24). Wang *et al* cometieron uno de los errores metodológicos más frecuentes. En este artículo se consideraron diferencias intragrupal entre remdesivir y placebo en el subgrupo de tratamiento temprano para el tiempo hasta la mejoría clínica y carga viral en el tracto respiratorio superior en el corte de 10 días de sintomatología (9,10,12,31). No se calculó el valor de $p(i)$ con el subgrupo complementario. De esta manera, se resaltaron diferencias numéricas para el tiempo hasta la mejoría clínica aparentemente favorables para el uso temprano de remdesivir, promoviendo una interpretación inadecuada del análisis de subgrupos que puede influir en la toma de decisiones clínicas. Además, no se deben aplicar siempre las diferencias de efecto encontradas en un análisis de subgrupos. Antes de ello, es necesario evaluar la relevancia clínica de las variables involucradas. Tanto la variable de carga viral como la de tiempo hasta mejoría clínica pueden considerarse medidas de escasa relevancia clínica si no se asocian con una reducción de la mortalidad.

Esta revisión encontró pacientes con diferentes características en los ECA sobre el uso de remdesivir en COVID-19 (31,92–95). Se seleccionaron estudios con COVID-19 moderada o severa y otros que reclutaron a ambas poblaciones. También se observó variabilidad en la oxigenación y ventilación de pacientes con COVID-19 en ECA. Beigel *et*

al y Kalil *et al* presentaron los peores resultados de la puntuación basal de las escalas ordinales sobre el uso de oxígeno en los pacientes (93,94). A esto hay que añadir que, además, los pacientes fueron evaluados utilizando diferentes escalas. Por otra parte, Beigel *et al* y Kalil *et al* desarrollaron los análisis por subgrupos según el momento de uso de remdesivir con los tamaños muestrales más grandes. En general, los estudios no proporcionaron toda la información esencial para el abordaje de datos de los subgrupos.

Este trabajo es un ejemplo ilustrativo de cómo un análisis metodológico de análisis por subgrupos puede evitar la realización de prematuras –y arriesgadas– afirmaciones. Se utilizaron dos publicaciones previas. Sun *et al* estableció una base adecuada para la consideración del análisis por subgrupos (4). Sin embargo, los evaluadores inexpertos pueden dudar sobre la importancia u orden de los criterios de interpretación. Esta limitación puede ser minimizada por la herramienta validada de Gil-Sierra *et al* (55), la cual valora otros criterios adicionales respecto a la primera metodología.

Aunque muchos medicamentos –como lopinavir/ritonavir o hidroxiclороquina– han sido ampliamente utilizados, hasta el momento de realización de esta investigación, sólo glucocorticoides y las vacunas habían mostrado una clara reducción de la mortalidad en COVID-19 (85,86,89,99). La incertidumbre acerca del efecto de las alternativas terapéuticas había sido muy elevado (87). El desarrollo de metodologías sistemáticas para evaluar nueva evidencia científica es necesario para reducir los efectos superfluos o negativos de los fármacos –y gastos innecesarios– en pacientes con COVID-19. Los costes mínimos de producción de remdesivir han sido estimados en 0.93\$/día (100). Sin embargo, el precio de adquisición de remdesivir es muy superior a los costes de producción. Por lo tanto, la optimización del uso de este antiviral es muy importante para la eficiencia de los sistemas de salud debido a su importante impacto económico.

Para finalizar, podemos decir que hemos realizado un estudio con una búsqueda sistemática y aplicación de una metodología establecida para la interpretación de análisis por subgrupos sobre el uso temprano de remdesivir en COVID-19. Además, nuestra revisión detalló estimaciones esenciales para la interpretación de análisis por

subgrupos no descritos anteriormente. Este trabajo encontró un beneficio estadísticamente significativo para el uso temprano de remdesivir en pacientes con COVID-19 severa y <7 días de síntomas, de acuerdo a las variables de tiempo hasta mejoría clínica y mejor escala de estado clínico en el día 15. Ningún otro beneficio fue asociado con el uso temprano de remdesivir en otras variables o cortes de tiempo. Por tanto, parece razonable la inclusión del corte de 7 días de inicio de los síntomas para evaluar el uso temprano de remdesivir en futuros estudios de COVID-19 severa.

3.3. Aplicación de la metodología y conocimientos desarrollados sobre publicaciones concretas.

3.3. APLICACIÓN DE LA METODOLOGÍA Y CONOCIMIENTOS DESARROLLADOS SOBRE PUBLICACIONES CONCRETAS.

Los errores de interpretación de los análisis por subgrupos suelen ser frecuentes en los ECA (6,21), a pesar de su notable impacto clínico en los pacientes y económico en los sistemas sanitarios. Uno de estos errores es la aceptación de heterogeneidad entre subgrupos complementarios cuando realmente no se han observado diferencias estadísticamente significativas. Esta situación se da cuando $p(i) > 0.1$. Otro error suele ser la aplicación de una diferencia indicativa [$p(i) < 0.1$, pero $p(i) > 0.05$] observada en análisis por subgrupos que carecen de una serie de características relevantes para favorecer premeditadamente el uso de una intervención sanitaria (4). Ejemplos de estas características pueden ser la pre-especificación en el protocolo de los ECA que valora la posible influencia de un factor en el beneficio obtenido de un fármaco, o la limitación de un número elevado de análisis de factores. De esta manera, se evitan interpretaciones *post hoc* –o *ad hoc*– y la aceptación de diferencias estadísticamente significativas generadas por un exceso de la multiplicidad de determinaciones. En otras ocasiones, se realizan afirmaciones prematuras a partir de análisis por subgrupos con escaso tamaño muestral. Sin lugar a dudas, esto contribuye a incrementar el grado de incertidumbre de las decisiones clínicas basadas en análisis por subgrupos.

Estos errores de consideración de los resultados por subgrupos podrían afectar a la selección adecuada de los datos a la hora de desarrollar estudios; a la eficacia y seguridad de los tratamientos; al posicionamiento terapéutico de los fármacos y a la eficiencia de los sistemas sanitarios. En este capítulo llevaremos a cabo la aplicación de la herramienta validada que hemos desarrollado sobre la interpretación de análisis por subgrupos en publicaciones concretas de ECA que realizan afirmaciones inapropiadas (55), los cuales han ido surgiendo en la práctica clínica habitual diaria del doctorando en el período de tiempo que ha realizado esta Tesis Doctoral. Los fármacos involucrados en este capítulo son pembrolizumab (101), abemaciclib (102), la combinación de nivolumab más ipilimumab (N+I) (103,104) y remdesivir (105).

3.3.1. Pembrolizumab más quimioterapia en cáncer de pulmón no microcítico (CPNM) metastásico.

Gil-Sierra MD, Fenix-Caballero S, Alegre-del Rey E. Pembrolizumab plus chemotherapy in Lung Cancer. *Engl J Med.* 2018;379(11):e18. Doi: 10.1056/NEJMc1808567.

3.3.1.1. Introducción

En el ensayo KEYNOTE-189, Gandhi *et al.* compararon la combinación de pembrolizumab más quimioterapia frente a quimioterapia en pacientes con CPNM metastásico no tratado previamente (101). En este estudio se observó un beneficio estadísticamente significativo en SG y SLP al añadir pembrolizumab a la quimioterapia en este escenario clínico.

PD-L1 es un biomarcador que ha adquirido una enorme importancia en los últimos años en la selección de medicamentos. La expresión de PD-L1 ha sido evaluada como criterio para la selección de fármacos en numerosos contextos clínicos (15,106,107), llegando a tener un papel considerable en los Informes de Posicionamiento Terapéutico (IPT). El análisis por subgrupos ha sido la forma más frecuente de valorar la relación entre el efecto de la inmunoterapia y los niveles de expresión de PD-L1. Por tanto, un abordaje sistemático y objetivo de los análisis por subgrupos de este biomarcador es esencial en el posicionamiento terapéutico de los nuevos tratamientos. Los autores del estudio KEYNOTE-189 llegaron a la conclusión de que los pacientes con CPNM no tratado previamente y una expresión de PD-L1 igual o superior al 50% obtuvieron un mayor beneficio con el esquema que combina inmunoterapia y quimioterapia (101). No obstante, existe un importante solapamiento entre el intervalo de confianza del subgrupo de pacientes con una expresión de PD-L1 de 1-49% y el intervalo de confianza del subgrupo que tiene una expresión igual o más del 50% (**Figura 9**). Asimismo, Gandhi *et al.* no detallaron el valor de p(i) entre estos subgrupos. Como ya hemos comentado, este dato es esencial para la toma de decisiones clínicas a partir de análisis por subgrupos.

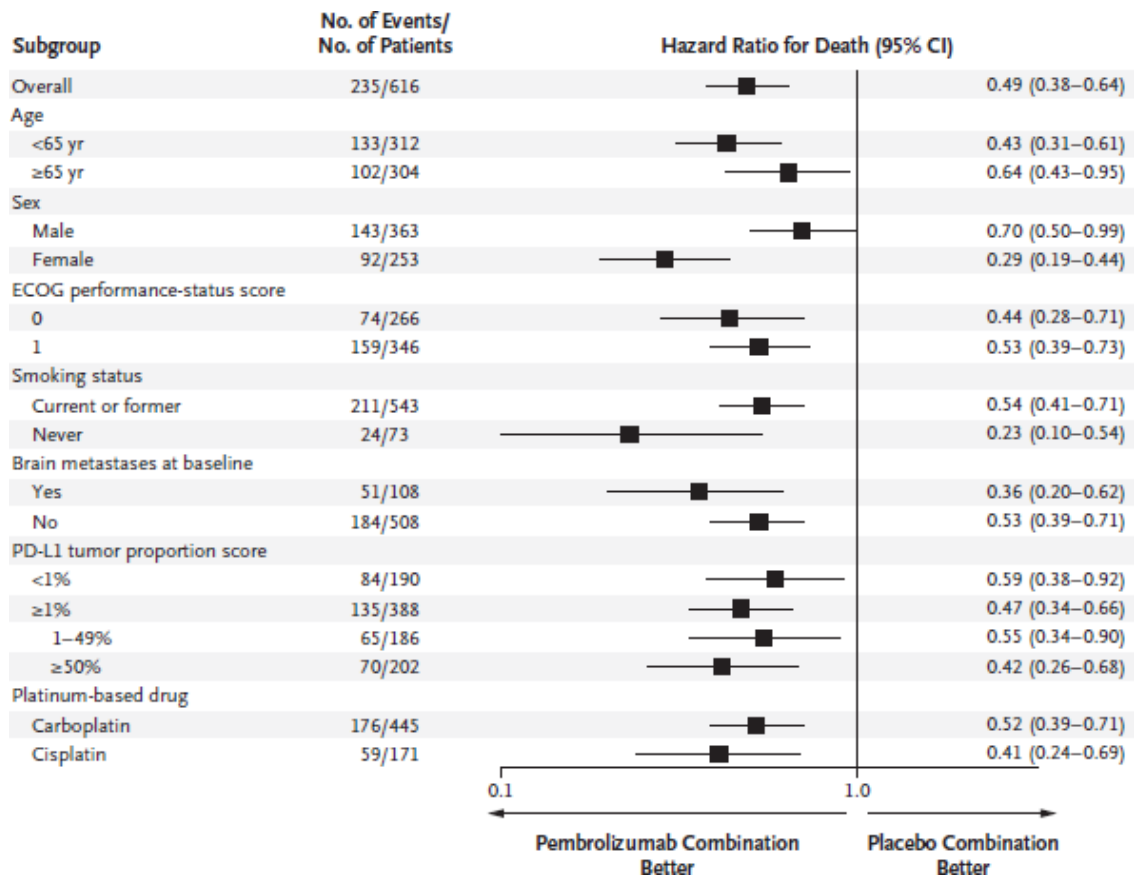


Figura 9. Análisis por subgrupos de SG del estudio KEYNOTE-189. Tomada de Gandhi *et al* (2018) (101)

Teniendo en cuenta todo lo anterior y las limitaciones del análisis por subgrupos (5), parece razonable que se llevara a cabo la valoración con nuestra herramienta de los datos de eficacia de pembrolizumab más quimioterapia en CPNM metastásico no tratado en los subgrupos citados según la expresión de PD-L1.

3.3.1.2. Metodología

Se usó la herramienta validada que hemos desarrollado para valorar la aplicabilidad del análisis por subgrupos (55). La herramienta se compone por unas cuestiones preliminares para descartar análisis por subgrupos sin condiciones mínimas relevantes, y un cuestionario. Una respuesta negativa en alguna de las cuestiones preliminares descartó la aplicabilidad del análisis por subgrupos de forma directa. El cuestionario

asoció valoraciones con puntuaciones para cada criterio de interpretación del análisis por subgrupos: asociación estadística, plausibilidad biológica y consistencia de resultados. Estas valoraciones y puntuaciones fueron: probable (+3 puntos), posible (+2 puntos), dudosa (0 puntos) y nula (-3 puntos). El sumatorio de las puntuaciones se relacionó con recomendaciones para la aplicabilidad en la práctica clínica: 9-7 puntos (recomendación “probable”) se asignó con aplicabilidad de resultados de subgrupos; 5-6 puntos (recomendación “posible”) con aplicación cautelosa ante dificultad de uso, escasa tolerancia o coste excesivo de alternativas terapéuticas; 3-4 puntos (recomendación “dudosa”) con evitar la aplicación de datos de subgrupos salvo excepciones; y <3 puntos (recomendación “nula”) con no aplicación de resultados de subgrupos. Siempre que la asociación estadística o consistencia fueron nulas, la aplicabilidad de los resultados de subgrupos fue descartada directamente, y se llevó a cabo una “recomendación nula”.

Para calcular el valor de $p(i)$ entre subgrupos –ya que el estudio no aportó su estimación– se usó una calculadora basada en la escala de Ln de los valores de riesgo relativo, odds ratio, hazard ratio e intervalos de confianza proporcionados por el ensayo clínico (12,33,68–72).

3.3.1.3. Resultados

Se realizaron estimaciones de $p(i)$. Para SG, el valor de $p(i)$ entre el subgrupo de pacientes con una expresión de PD-L1 <1% y el subgrupo con expresión $\geq 1\%$ fue 0,42. Asimismo, $p(i)$ entre el subgrupo de expresión de PD-L1 con valor de 1-49% y el subgrupo con $\geq 50\%$ fue 0,44. El valor de $p(i)$ entre los subgrupos con expresión de PD-L1 <1%, 1-49% y $\geq 50\%$ fue 0,57. Posteriormente, se aplicó la herramienta de interpretación de análisis por subgrupos. La pregunta preliminar sobre existencia de diferencias de efecto entre subgrupos recibió respuesta negativa [$p(i) > 0.1$]. Por tanto, no se aplicó el resto de la herramienta y se descartó la aplicabilidad del análisis por subgrupos (**Tabla 10**).

Tabla 10. Resumen de evaluación sobre análisis por subgrupos según la expresión de PD-L1

Metodología	Criterios		Gandhi <i>et al.</i> (ensayo KEYNOTE-189), evaluación de SG†
Herramienta validada (Gil-Sierra <i>et al.</i>) (55)	Cuestiones preliminares	Nivel más alto de evidencia del estudio con análisis por subgrupos	Sí
		Relevancia clínica clara de variable considerada o variable primaria subrogada	Sí
		Existencia de diferencia en el efecto entre los subgrupos para el factor analizado [$p(i) < 0.1$]	No
		Factor determinante del análisis por subgrupos previo a la intervención sanitaria	No aplicado
	Cuestionario Asociación estadística (puntuación) Plausibilidad biológica (puntuación) Consistencia (puntuación) Recomendación	No aplicado No aplicado No aplicado No aplicado No aplicado	

†SG: supervivencia global

3.3.1.4. Discusión

Los resultados obtenidos al aplicar nuestra herramienta sobre interpretación de análisis por subgrupos según el grado de expresión para PD-L1 para la variable SG no permiten aceptar la afirmación de los autores del ensayo KEYNOTE-189 (101), en la que se asocia un mayor beneficio de pembrolizumab más quimioterapia en pacientes con una expresión de PD-L1 igual o superior al 50%. La ausencia de heterogeneidad entre el subgrupo con menos del 50% de expresión de PD-L1 y el subgrupo con una expresión de igual o más del 50% no permite asumir las diferencias sugeridas. Gandhi *et al* llegaron a conclusiones sobre estos subgrupos de pacientes sin describir $p(i)$. El cálculo de la interacción estadística es esencial para aceptar diferencias sugeridas en los análisis por subgrupos e incrementar la credibilidad de estos datos.

Aunque en el estudio KEYNOTE-189 se muestran análisis por subgrupos según la expresión de PD-L1 para SLP, las estimaciones de $p(i)$ entre los subgrupos para esta variable tienen menos importancia para la toma de decisiones clínicas que los datos de SG descritos. Si bien es cierto que ambas son variables irreversibles de importante impacto clínico, SG es una variable final y SLP es una variable subrogada. La selección y evaluación de medicamentos siempre debe considerar en mayor medida aquellas variables que claramente mejoren la duración y calidad de vida de los pacientes frente a otras que pueden estar (o no) relacionadas con estos beneficios. De hecho, existen estudios que señalan que los propios pacientes lo que realmente quieren es, obviamente, vivir más y mejor (58,59,61). SLP es una variable que puede estar relacionada con SG, pero su medida puede estar influenciada por el centro seleccionado para valorar las variables, profesional evaluador o los criterios establecidos como progresión de enfermedad. De hecho, SLP no está exenta de controversia en muchos casos y se ha señalado que SG debería ser la variable primaria de los ensayos clínicos de cáncer de pulmón (108). Por este motivo, no se ha llevado a cabo la aplicación de nuestra herramienta en los datos de los subgrupos para SLP, ya que se dispone de información de SG. No obstante, el estudio KEYNOTE-189 muestra que el solapamiento es muy elevado entre los *forest plot* que representan los valores de hazard ratio e ICs para SLP

de los subgrupos con expresión de PD-L1 del 1-49% y $\geq 50\%$ (101). De hecho, puede afirmarse que el valor de $p(i)$ es superior a 0,1 entre estos subgrupos. Por tanto, los valores de SLP también respaldarían una ausencia de mayor beneficio de la combinación de pembrolizumab más quimioterapia en pacientes con una expresión de PD-L1 igual o mayor al 50%.

Asumir diferencias no reales entre subgrupos puede tener consecuencias negativas en la selección de datos para futuros estudios. Por ejemplo, si se acepta un valor de eficacia de un tratamiento en un subgrupo de pacientes de un ensayo clínico, se podría usar para el desarrollo de futuras comparaciones indirectas o network meta-análisis (NMA) de terapias en un determinado contexto clínico. Si la interpretación del análisis por subgrupos no fue adecuada y no debería haberse asumido el valor de ese subgrupo de pacientes, esa comparación indirecta o NMA posterior podría llegar a conclusiones erróneas. Quizás en comparaciones indirectas o NMA que incluyan el esquema de pembrolizumab más quimioterapia en pacientes con CPNM metastásico no tratado previamente y expresión de PD-L1 igual o mayor al 50% lo más aconsejable sería realizar análisis de sensibilidad con datos de subgrupos y los resultados globales del ensayo. Por otra parte, la aplicación inadecuada de resultados de subgrupos de pacientes también puede afectar a la eficacia, seguridad y posicionamiento terapéutico de los tratamientos.

Los pacientes con CPNM metastásico no tratado previamente y expresión de PD-L1 igual o mayor al 50% pueden ser tratados con pembrolizumab en monoterapia (17). Teniendo en cuenta que Gandhi *et al* afirmaron que pembrolizumab más quimioterapia obtuvo el mayor beneficio en estos pacientes, cabría preguntarse cuál de los dos esquemas debería ser priorizado. Existen comparaciones indirectas ajustadas que han estimado que no existen diferencias estadísticamente significativas para SG entre la combinación de pembrolizumab más quimioterapia y pembrolizumab en monoterapia (109,110) (**Tabla 11**). Estos estudios afirman que existe un dudoso beneficio de pembrolizumab más quimioterapia frente a pembrolizumab en monoterapia en pacientes con CPNM metastásico no tratado previamente y expresión de PD-L1 igual o mayor al 50%, y que la combinación debe ser considerada con cautela. La combinación de pembrolizumab más quimioterapia presenta peor perfil de seguridad y mayor coste que pembrolizumab en

monoterapia. Pembrolizumab en monoterapia permite reservar la quimioterapia basada en dobletes de platino para líneas posteriores. De esta manera, los pacientes tendrían una opción terapéutica más ante una hipotética progresión tumoral a la inmunoterapia. Teniendo en cuenta todo lo anterior, parece razonable priorizar el uso del esquema de pembrolizumab en monoterapia en pacientes con CPNM metastásico no tratado previamente y expresión de PD-L1 igual o mayor al 50%, en lugar de la combinación de pembrolizumab más quimioterapia.

Tabla 11. Datos de comparación indirecta (CI) de pembrolizumab más quimioterapia (Pb-CT) frente a pembrolizumab en monoterapia. Tomado de Gil-Sierra *et al.* Indirect comparison of pembrolizumab plus chemotherapy versus pembrolizumab in lung cancer (2019) (109).

Reference	PFS	OS
Pb-CT vs. CT	HR=0.36 (95% CI, 0.25 to 0.52, PD-L1 ≥50% subgroup)	HR=0.42 (95% CI, 0.26 to 0.68, PD-L1 ≥50% subgroup).
Pb vs. CT	HR 0.63 (95% CI, 0.44 to 0.91, subgroup platinum+pemetrexed).	HR 0.60 (95% CI, 0.41 to 0.89)
Pb-CT vs. Pb (ITC)	HR=0.57 (95% CI, 0.40 to 0.96)	HR=0.70 (95% CI, 0.38 to 1.30)

Este es un claro ejemplo de como una afirmación sobre subgrupos llevada a cabo por Gandhi *et al* (101) puede afectar negativamente a la selección de datos para el desarrollo de estudios, eficacia, seguridad, eficiencia y posicionamiento terapéutico de fármacos en un determinado contexto clínico. Si se hubiera aceptado sin más que pembrolizumab mas quimioterapia obtiene el mejor resultado en pacientes con CPNM metastásico no tratado previamente y expresión de PD-L1 igual o mayor al 50%, probablemente muchos clínicos estarían tratando a sus pacientes con este esquema. Los pacientes y el sistema sanitario se habrían visto afectados negativamente porque pembrolizumab más quimioterapia no ha mostrado una mejor eficacia que pembrolizumab en monoterapia, presenta peor perfil de seguridad, menor eficiencia y no permite reservar la quimioterapia basada en dobletes de platino para líneas de tratamiento posteriores.

3.3.2. Abemaciclib más fulvestrant en cáncer de mama con receptor hormonal positivo y ERBB2 negativo tras progresión a terapia endocrina.

Gil-Sierra MD, Briceño-Casado MDP, Fénix-Caballero S. Analysis of overall survival benefit of abemaciclib plus fulvestrant in hormone receptor-positive, ERBB2-Negative Breast Cancer. *JAMA Oncol.* 2020;6(7):1122. Doi: 10.1001/jamaoncol.2020.1516.

3.3.2.1. Introducción

En el estudio MONARCH 2, Sledge *et al.* compararon abemaciclib asociado a fulvestrant frente a placebo más fulvestrant para el tratamiento de cáncer de mama avanzado con receptor hormonal positivo y ERBB2 negativo tras progresión a terapia endocrina (102). Este ensayo encontró un beneficio estadísticamente significativo para SG al añadir abemaciclib a fulvestrant en este contexto clínico.

El análisis por subgrupos presenta un carácter exploratorio con el fin de detectar diferencias de efecto de una intervención sanitaria en diferencias subpoblaciones según características concretas. De esta manera, puede servir como generador de hipótesis o herramienta que pudiera ayudar a la optimización del posicionamiento terapéutico. No obstante, no hay que olvidar las limitaciones del análisis por subgrupos –como el incremento de los errores α y β – durante la toma de decisiones clínicas (5). Los autores del estudio MONARCH 2 afirmaron que el mayor beneficio fue observado en los subgrupos de pacientes con metástasis viscerales y resistencia primaria a la terapia endocrina (102). Estos comentarios se basaron en la separación temprana de las curvas en las gráficas de SG y en diferencias numéricas (**Figura 10**). La consideración de diferencias intragrupalas o separación de las curvas en resultados de subgrupos no es recomendable, contrariamente a lo que se ha interpretado en el estudio MONARCH 2. Las aparentes diferencias de efecto pueden deberse al azar o a un desbalanceamiento de la distribución de factores relacionados con el beneficio. Este tipo de hallazgos podría verse favorecido por la redistribución de la población del estudio en subpoblaciones más pequeñas y un aumento de la incertidumbre en el análisis por subgrupos.

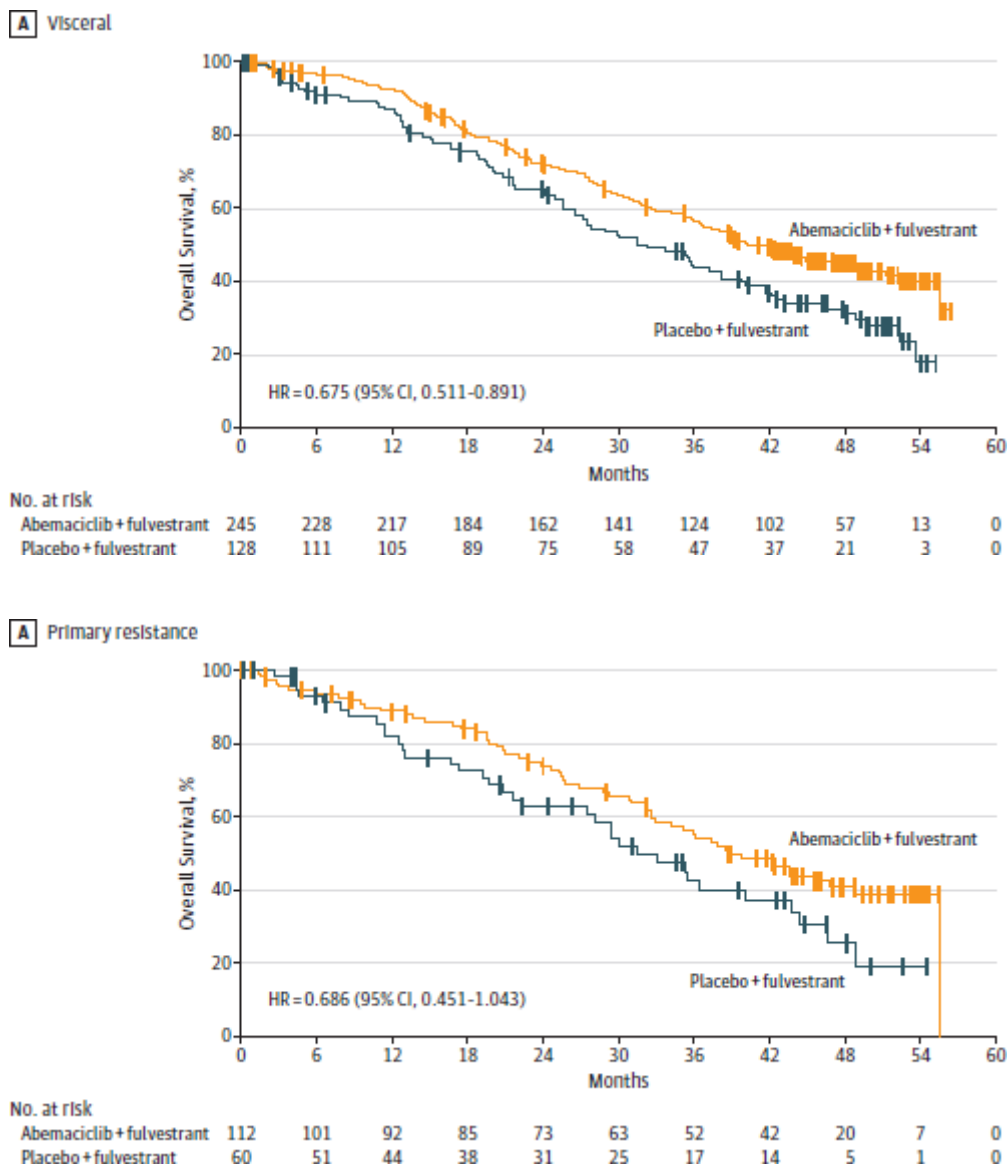


Figura 10. Representaciones gráficas de las curvas de SG de los subgrupos de pacientes con metástasis viscerales y resistencia primaria a la terapia endocrina (diferencias intragrupalas). Tomada de Sledge *et al* (2020) (102).

Por otra parte, el artículo describe que no se encontró heterogeneidad entre estos subgrupos y sus complementarios, sin detallar el valor de $p(i)$. La representación del *forest plot* del análisis por subgrupos del estudio MONARCH 2 muestra un importante solapamiento entre los intervalos de confianza de los subgrupos de pacientes con metástasis viscerales y resistencia primaria a la terapia endocrina con sus respectivos

subgrupos complementarios (**Figura 11**). Como ya hemos comentado anteriormente, la estimación de la heterogeneidad entre subgrupos es esencial para valorar la fiabilidad de una aparente diferencia de efecto en un análisis por subgrupos (4,55).

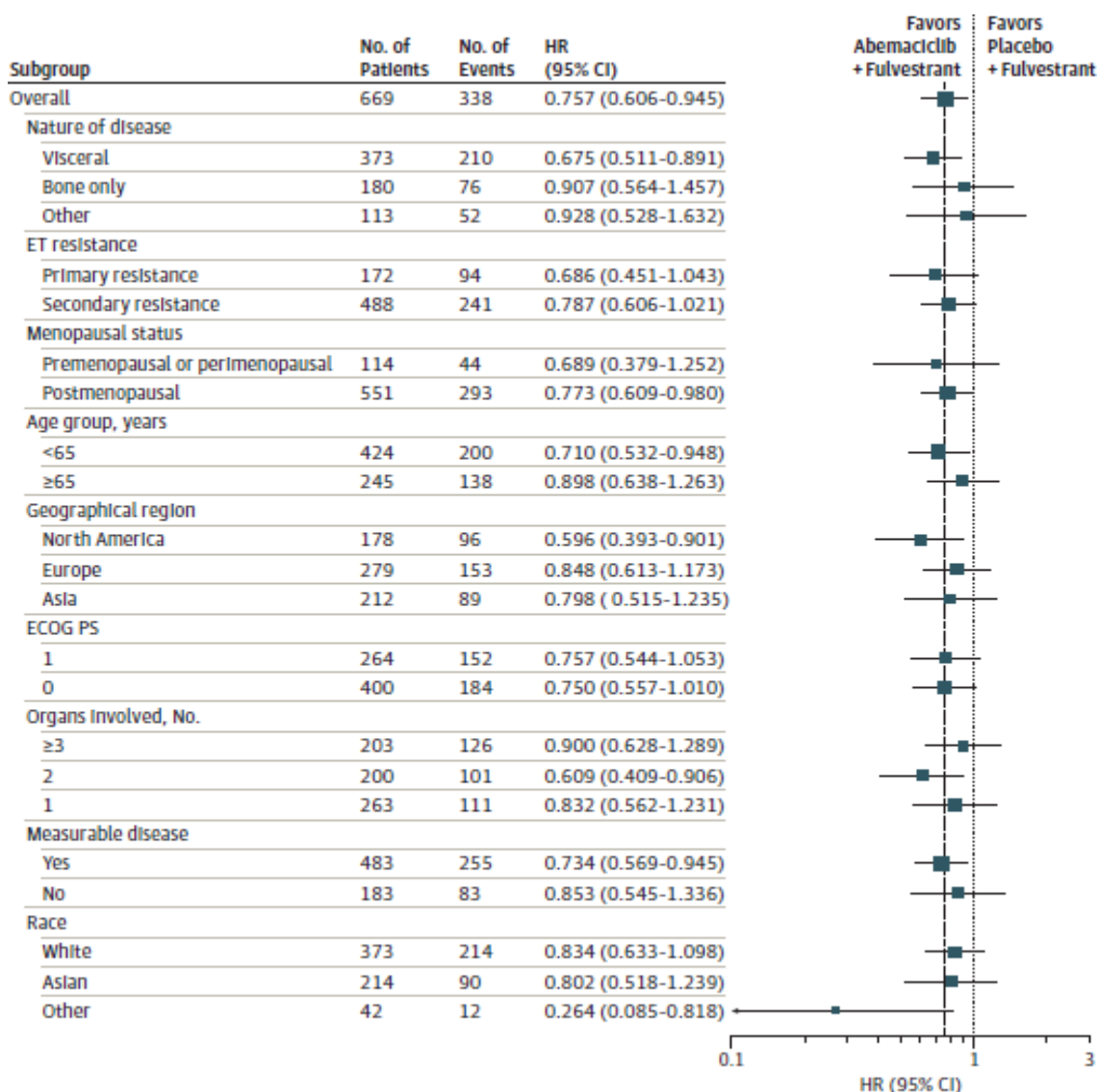


Figura 11. Análisis por subgrupos de SG del estudio MONARCH 2. Tomada de Sledge *et al* (2020) (102)

Teniendo en cuenta todo lo anterior, ¿podemos aceptar que la combinación de abemaciclib con fulvestrant obtuvo los mejores resultados en los subgrupos con metástasis viscerales y resistencia primaria a la terapia endocrina en cáncer de mama

avanzado con receptor hormonal positivo y ERBB2 negativo? A continuación, aplicaremos la herramienta desarrollada sobre interpretación de análisis por subgrupos para responder a esta cuestión.

3.3.2.2. Metodología

Se usó nuestra herramienta validada para valorar la aplicabilidad del análisis por subgrupos (55). Esta herramienta se estructura en dos partes: cuestiones preliminares para descartar la valoración de los análisis por subgrupos sin condiciones mínimas relevantes, y un cuestionario. Estas cuestiones preliminares valoran el nivel de evidencia del estudio que contiene el análisis por subgrupos, relevancia de la variable seleccionada, existencia de diferencia de efecto [$p(i) < 0.1$] entre subgrupos para el factor seleccionado, y secuencia temporal entre este factor y la intervención sanitaria. La respuesta negativa de alguna cuestión preliminar conlleva al descarte de la realización de las preguntas posteriores, del cuestionario y aplicabilidad del análisis por subgrupos. El cuestionario asocia una valoración y puntuación a cada uno de los criterios de la interpretación del análisis por subgrupos: asociación estadística, que incluye $p(i)$, pre-especificación de subgrupos, tamaño muestral, número de factores estudiados y resultado global del estudio; plausibilidad biológica y consistencia de los resultados. La puntuación es más elevada cuanto mayor sea la fiabilidad del criterio: probable (+3 puntos), posible (+2 puntos), dudosa (0 puntos) y nula (-3 puntos). El sumatorio global de las puntuaciones de los criterios se corresponde con una recomendación para la aplicabilidad de los resultados en la práctica clínica. Las puntuaciones más altas del cuestionario se asocian con una mayor aplicabilidad de los datos del análisis por subgrupos y, por consiguiente, con recomendaciones para que deban tenerse en consideración para la toma de decisiones clínicas. Los sumatorios de las puntuaciones son relacionados con las siguientes recomendaciones: 9-7 puntos (recomendación “probable”) se corresponde con aplicabilidad de los resultados por subgrupos; 5-6 puntos (recomendación “posible”) se asocia con aplicabilidad prudente ante escasa tolerancia, dificultad de uso o coste de otras intervenciones; 3-4 puntos (recomendación “dudosa”) se relaciona con desechar la aplicabilidad del análisis por subgrupos salvo

excepciones; y <3 puntos (recomendación “nula”) se relaciona con la no aplicación de los datos de los subgrupos.

Para calcular los valores de $p(i)$ entre los subgrupos –datos no aportados por el ensayo– se usaron calculadoras basadas en la escala del Ln de los valores de riesgo relativo, odds ratio, hazard ratio e intervalos de confianza disponibles (12,33,68–72).

3.3.2.3. Resultados

Se realizaron estimaciones de $p(i)$. Para SG, el análisis por subgrupos según la naturaleza de la enfermedad dividió a la población global en tres subgrupos: metástasis viscerales, enfermedad exclusivamente en huesos y “otros”. $P(i)$ presentó un valor de 0,29 entre el subgrupo con enfermedad visceral y el subgrupo de pacientes con enfermedad exclusivamente en huesos. $P(i)$ fue 0,32 entre el subgrupo con enfermedad visceral y el subgrupo “otros”. No se encontró heterogeneidad entre los tres subgrupos [$p(i) = 0,42$]. Teniendo en cuenta la variable SG, el análisis por subgrupos según la resistencia a la terapia endocrina dividió a la población global en dos subgrupos: resistencia primaria y resistencia secundaria. El valor de $p(i)$ entre estos dos subgrupos fue 0,585.

Tras esto, se usó la herramienta de aplicación de análisis por subgrupos. Se respondió negativamente a la pregunta preliminar sobre existencia de diferencias de efecto entre subgrupos [$p(i) > 0.1$]. Por tanto, no se siguió aplicando el resto de la herramienta y se desechó la aplicabilidad del análisis por subgrupos (**Tabla 12**).

Tabla 12. Resumen de evaluación sobre análisis por subgrupos

Metodología	Criterios		Sledge <i>et al</i> (estudio MONARCH 2), evaluación de SG [†]	
			Factores del análisis por subgrupos: subgrupo seleccionado	
			Naturaleza de la enfermedad: subgrupo con metástasis viscerales	Resistencia a terapia endocrina: subgrupo con resistencia primaria
Herramienta validada (Gil-Sierra <i>et al.</i>) (55)	Cuestiones preliminares	Nivel más alto de evidencia del estudio con análisis por subgrupos	Sí	Sí
		Relevancia clínica clara de variable seleccionada o variable primaria subrogada	Sí	Sí
		Diferencia de efecto entre subgrupos para factor considerado [$p(i) < 0.1$]	No	No
		Factor determinante del análisis por subgrupos previo a la intervención sanitaria	No aplicado	No aplicado
	Checklist Asociación estadística (puntuación) Plausibilidad biológica (puntuación) Consistencia (puntuación) Recomendación	No aplicado	No aplicado	

†SG: supervivencia global

3.3.2.4. Discusión

La recomendación obtenida por nuestra herramienta de aplicabilidad de análisis por subgrupos en cáncer de mama avanzado con receptor hormonal positivo y ERBB2 negativo tras progresión a terapia endocrina no invita a aceptar diferencias de efecto de abemaciclib entre los subgrupos con metástasis viscerales y resistencia primaria a la terapia endocrina con las respectivas subpoblaciones complementarias (55,102). Las afirmaciones de una aparente diferencia de efecto de este fármaco en estos dos subgrupos se basan en una temprana separación de las curvas de SG y diferencias de efecto numéricas. La consideración de representaciones gráficas de SG de análisis por subgrupos sin valorar la heterogeneidad a través del valor de $p(i)$ no es una evaluación metodológicamente aceptable. La valoración de diferencias numéricas no estadísticamente significativas tampoco constituye un abordaje recomendable. Estas intervenciones podrían dar lugar a conclusiones engañosas.

En el estudio MONARCH 2 se aceptó como beneficio asociado a abemaciclib en determinados subgrupos a una diferencia numérica no estadísticamente significativa. Este tipo de errores metodológicos podría desencadenar el encarecimiento de una determinada alternativa terapéutica con un “aparente beneficio” en un escenario clínico en el que otros fármacos competidores no lo han conseguido, cuando en realidad el beneficio es inexistente o no se ha podido demostrar de forma consistente. De esta manera, se perjudica la competitividad entre fármacos y la adquisición de medicamentos a un precio menor.

A la vista de estas consideraciones, no puede concluirse que abemaciclib más fulvestrant obtuvo los mejores resultados en los subgrupos de pacientes con metástasis viscerales y resistencia primaria a la terapia endocrina en cáncer de mama avanzado con receptor hormonal positivo y ERBB2 negativo tratado previamente con terapia endocrina. Por tanto, parece razonable valorar resultados más maduros de SG y ser cauteloso con los datos de subgrupos del estudio MONARCH 2 (102).

3.3.3. Nivolumab más ipilimumab en carcinoma escamoso de la cavidad oral no tratado previamente.

Gil-Sierra MD, Sánchez-Hidalgo M, Alarcón de la Lastra-Romero C. Clinical decision-making about neoadjuvant nivolumab plus ipilimumab. *JAMA Oncol.* 2021;7(2):309. Doi: 10.1001/jamaoncol.2020.6989

3.3.3.1. Introducción

En 2020, Schoenfeld *et al* desarrollaron un ECA abierto de fase II para comparar el esquema de N+I frente a nivolumab como tratamiento neoadyuvante en pacientes con carcinoma de células escamosas de la cavidad oral no tratado previamente (103). A pesar de las limitaciones del estudio, los autores de este artículo presentaron a la combinación de N+I como un tratamiento neoadyuvante factible en estos pacientes.

Se reclutaron tan sólo 29 pacientes en este ECA. Para medir la eficacia de las terapias, se midieron variables de dudosa relevancia clínica que no han sido validadas (respuestas volumétrica y patológica), y otras variables finales, como SG y SLP. Sin embargo, estas dos últimas presentaron datos inmaduros. Para valorar la tolerancia de los tratamientos, se seleccionó como variable principal a los retrasos de la fecha quirúrgica planificada y también se registraron los efectos tóxicos limitantes de dosis. No se observaron ni retrasos de la fecha quirúrgica planificada ni efectos tóxicos limitantes de dosis.

Por otra parte, se evaluó la relación entre potenciales biomarcadores predictores de respuesta a la terapia (como PD-L1) con los resultados obtenidos por los tratamientos. La asociación entre la presencia de estos biomarcadores y la respuesta a los fármacos se determinó utilizando *rank sum test*, regresión logística univariada y análisis de varianza. No se encontró correlación entre la expresión de PD-L1 con la respuesta volumétrica y patológica de la cohorte global ni en el brazo de nivolumab en monoterapia. Contrariamente, se asoció un grado de respuesta patológica al tratamiento con la

presencia de células T positivas para el cúmulo de diferenciación 4 (CD4, del inglés *cluster of quadruple differentiation*) en la población global, y se afirmó que esta asociación fue significativa en la cohorte de pacientes tratados con N+I, pero no para el brazo de nivolumab en monoterapia.

El análisis por subgrupos tiene un reconocido papel como generador de nuevas hipótesis ante determinados hallazgos. No obstante, teniendo en cuenta las limitaciones del análisis por subgrupos y del trabajo de Schoenfeld *et al* (4,103), ¿pueden ser tenidos en cuenta los resultados observados en los citados biomarcadores? Para responder a esta pregunta, hemos aplicado nuestra herramienta validada al efecto de nivolumab N+I como tratamiento neoadyuvante en pacientes con carcinoma de células escamosas de la cavidad oral no tratado previamente y los biomarcadores PD-L1 y CD4.

3.3.3.2. Metodología

Se empleó nuestra herramienta de valorar la aplicabilidad del análisis por subgrupos (55), la cual se divide en dos partes: cuestiones preliminares para descartar los análisis por subgrupos sin condiciones mínimas relevantes, y un cuestionario. Las preguntas preliminares valoran el nivel de evidencia del estudio, relevancia clínica de la variable evaluada, diferencia de efecto entre los subgrupos [$p(i) < 0.1$] y secuencia temporal entre el factor diferenciador de subpoblaciones y la intervención sanitaria. La respuesta de las cuestiones preliminares, desarrollo del *checklist* y aplicabilidad del análisis por subgrupos son descartados ante la respuesta negativa en una de las cuestiones preliminares. El *checklist* relaciona valoraciones con puntuaciones de cada criterio de interpretación del análisis por subgrupos: asociación estadística, que incluye $p(i)$, pre-especificación del análisis, tamaño muestral de subgrupos, número de factores considerados y resultado global del estudio; plausibilidad biológica y consistencia de resultados. Los rangos de las valoraciones y puntuaciones se detallan a continuación: probable (+3 puntos), posible (+2 puntos), dudosa (0 puntos) y nula (-3 puntos). El sumatorio global de las puntuaciones de los tres criterios (asociación estadística, plausibilidad biológica y consistencia) es relacionado con una recomendación para

aplicar los resultados de los subgrupos en la práctica clínica. Las puntuaciones más altas del cuestionario se corresponden con una mayor aplicabilidad de los resultados de los subgrupos: 9-7 puntos (recomendación “probable”) se asocia con aplicabilidad del análisis por subgrupos; 5-6 puntos (recomendación “posible”) con aplicabilidad cautelosa ante escasa tolerancia, dificultad de utilización o coste de otras intervenciones; 3-4 puntos (recomendación “dudosa”) se relaciona con evitar aplicabilidad de resultados de subgrupos salvo excepciones; y <3 puntos (recomendación “nula”) se relaciona con el rechazo de la aplicación del análisis por subgrupos. Ante una asociación estadística o consistencia “nulas”, la aplicabilidad del análisis se descartaría directamente sin valorar el sumatorio de los criterios (recomendación “nula”).

3.3.3.3. Resultados

Se aplicó la herramienta de interpretación de análisis por subgrupos a las afirmaciones sobre los siguientes biomarcadores: CD4 en células T y PD-L1. La hipotética ausencia de relación entre el efecto de N+I y la expresión de PD-L1 en el contexto clínico abordado fue evaluada mediante la respuesta volumétrica y patológica. La presunta relación entre la eficacia de N+I y la expresión de CD-4 en células T se valoró mediante la respuesta patológica. Cabe recordar que estas variables son de dudosa relevancia clínica, y que los resultados de SG y SLP del estudio son aún inmaduros.

Por tanto, se respondió negativamente a la pregunta preliminar sobre la relevancia clínica clara de variable seleccionada o variable primaria subrogada. Teniendo todo esto en cuenta, no se siguió aplicando el resto de la herramienta y se rechazó la aplicabilidad del análisis por subgrupos (**Tabla 13**).

Tabla 13. Resumen de evaluación sobre análisis por subgrupos

Metodología	Criterios		Schoenfeld <i>et al</i>	
			Factor del análisis por subgrupos evaluado, variable seleccionada	
			Expresión de PD-L1, respuesta volumétrica y patológica	Expresión de CD-4 en células T, respuesta patológica
Herramienta validada (Gil-Sierra <i>et al.</i>) (55)	Cuestiones preliminares	Nivel más alto de evidencia del estudio con análisis por subgrupos	Sí	Sí
		Relevancia clínica clara de variable seleccionada o variable primaria subrogada	No	No
		Diferencia de efecto entre subgrupos para factor considerado [$p(i) < 0.1$]	No	No
		Factor determinante del análisis por subgrupos previo a la intervención sanitaria	No aplicado	No aplicado
	Checklist Asociación estadística (puntuación) Plausibilidad biológica (puntuación) Consistencia (puntuación) Recomendación	No aplicado	No aplicado	

3.3.3.4. Discusión

Los resultados obtenidos por nuestra herramienta de interpretación de análisis por subgrupos no respaldan el descarte definitivo de la relación entre el efecto de N+I y la expresión de PD-L1 en el tratamiento neoadyuvante de pacientes con carcinoma de células escamosas de la cavidad oral no tratados previamente (103), con los datos con los que disponemos. Por otra parte, tampoco se puede aceptar de forma concluyente una asociación entre la eficacia de N+I y la expresión de CD-4 en células T. La aplicación de un análisis por subgrupos según estos biomarcadores fue rechazada por la falta de valoración de variables con relevancia clínica clara. Sin embargo, el diseño del estudio de Schoenfeld *et al* también podría ser considerado negativo en la primera pregunta preliminar de nuestra herramienta de subgrupos (55,103). Se trata de un ensayo clínico fase II con escaso tamaño muestral y sin datos maduros de variables importantes, por lo que podría no ser valorado como nivel más alto de evidencia. Para llegar a conclusiones con una fiabilidad aceptable es necesario que se desarrollen estudios fase III, o al menos, fase II con mayor tamaño muestral que presenten resultados maduros de variables validadas y robustas (como SLP y SG).

También es reseñable que en el trabajo de Schoenfeld *et al* se realicen afirmaciones sobre la falta de correlación entre la expresión de PD-L1 y la eficacia de los tratamientos (N+I y nivolumab) en la población global, y sobre el brazo de nivolumab en monoterapia, pero que no aporte ningún comentario sobre el brazo de N+I. De hecho, tenemos que recordar que las valoraciones intragrupalas de los datos de subgrupos no son recomendables (4), así como que no se estime el valor de $p(i)$ entre subgrupos complementarios según niveles de expresión de PD-L1. Por otra parte, aún no se puede aceptar una hipotética relación entre la eficacia de N+I y la expresión de CD-4 en células T. Son necesarios análisis por subgrupos pre-especificados que incluyan mayores tamaños muestrales, y que tengan estimaciones de $p(i)$ entre subgrupos complementarios en diferentes cortes de expresión de estos biomarcadores.

El análisis por subgrupos supone una buena herramienta para la generación de hipótesis y como punto de partida de nuevas investigaciones. No obstante, la realización de afirmaciones basadas en resultados de subgrupos conlleva un elevado grado de incertidumbre cuando no se evalúan criterios objetivos. En el caso de N+I como tratamiento neoadyuvante en pacientes con carcinoma de células escamosas de la cavidad oral no tratados previamente (103), además hay otras consideraciones importantes. Aunque no se observaron efectos adversos limitantes de dosis o retrasos quirúrgicos en la cohorte de N+I, se debe hacer un seguimiento de otros datos que podrían tener implicaciones clínicas significativas.

Se registraron 34 eventos adversos grado 3-4 en el brazo N+I. Se apreció una diferencia considerable en comparación con los 14 eventos adversos grado 3-4 registrados en el brazo de nivolumab, teniendo en cuenta el pequeño tamaño muestral del estudio. Los eventos adversos graves pueden requerir un manejo más difícil del paciente. Asimismo, la estancia hospitalaria fue numéricamente más larga con el esquema N+I que con nivolumab en monoterapia. El tratamiento de N+I también presentó un inicio de los efectos tóxicos numéricamente más temprano que el brazo control. Todo esto podría tener una influencia negativa en la calidad de vida de los pacientes. Por tanto, parece razonable revisar el perfil de seguridad de la combinación N+I en estudios con mayor tamaño muestral, teniendo en cuenta la escasa tolerancia que mostró en otras neoplasias (111).

Por otro lado, la eficiencia siempre debe jugar un papel fundamental durante la toma de decisiones clínicas. Los pacientes con carcinoma celular escamoso de cabeza y cuello reseca consumen importantes recursos sanitarios (112). Datos publicados previamente sobre el uso de combinaciones de fármacos en este escenario clínico sugieren un aumento de los costes asociados a estos pacientes. Es importante que se tenga en cuenta el uso de tratamientos biológicos de rescate, como el caso del paciente con colitis tratado con vedolizumab en el ensayo de Schoenfeld *et al* (103). Además de

estos detalles, no debe olvidarse que la carga económica de esta patología tiene relación con el estadio de la enfermedad.

Son necesarios futuros estudios sobre el uso de N+I como tratamiento neoadyuvante en pacientes con carcinoma de células escamosas de la cavidad oral no tratados previamente para encontrar un adecuado balance de beneficio-riesgo y eficiencia. La identificación de poblaciones con mayor beneficio podría llevarse a cabo a través de un análisis por subgrupos pre-especificado con estimaciones de $p(i)$ (4,55). La relación entre la respuesta a N+I y la expresión de biomarcadores (por ejemplo, PD-L1 y CD4 en células T) u otras características de los pacientes no deben ser rechazadas o aceptadas hasta que se lleven a cabo estudios más importantes con análisis de variables finales. Por último, sería aconsejable ser cauteloso en la toma de decisiones clínicas sobre el esquema N+I en neoadyuvancia de carcinoma de células escamosas de la cavidad oral no tratado previamente hasta que se disponga de datos más maduros sobre la relación beneficio-riesgo del tratamiento, eficacia en subgrupos, y eficiencia

3.3.4. Nivolumab más ipilimumab en pacientes con CPNM escamoso estadio IV tratado previamente

Gil-Sierra MD, Briceño-Casado MDP, Sánchez-Hidalgo M. Nivolumab plus ipilimumab for previously treated stage IV squamous cell lung cancer. *JAMA Oncol.* 2022;8(4):642. Doi: 10.1001/jamaoncol.2021.7787.

3.3.4.1. Introducción

El ensayo clínico fase III denominado Lung-MAP S1400I comparó la combinación de N+I frente a nivolumab en monoterapia en pacientes con CPNM escamoso avanzado tratado previamente –aunque no con inmunoterapia– (104). Este estudio reportó un resultado global negativo, ya que no encontró beneficio de SG asociado a N+I en comparación con nivolumab en este contexto clínico.

La importancia de los biomarcadores como predictores de respuesta al tratamiento de las neoplasias es cada vez mayor. La expresión de PD-L1 suele medirse frecuentemente en los ECA del uso de la inmunoterapia para valorar su efecto (15,101). Recientemente, también se está evaluando el impacto de otros biomarcadores emergentes en la eficacia de las terapias, como la carga de mutación tumoral (TMB, del inglés *tumor mutation burden*). El ensayo clínico Lung-MAP S1400I es un ejemplo (104). Este estudio desarrolló un análisis por subgrupos teniendo en cuenta la expresión de PD-L1 y los niveles de TMB. Los autores predijeron un aumento de SG asociado al uso de N+I en pacientes pre-tratados con CPNM escamoso avanzado, y con expresión de PD-L1 menor al 1% y una TMB igual o mayor a 10 mutaciones por megabase (mt/Mb).

Como ya hemos citado en varias ocasiones, el análisis por subgrupos puede ser una herramienta interesante como generador de nuevas hipótesis o determinación de nuevos hallazgos en subpoblaciones. Sin embargo, presenta una serie de limitaciones (2,4,5). Los resultados de los subgrupos del ensayo Lung-MAP S1400I se obtuvieron a través de un análisis *post hoc* de un escaso número de pacientes. A esto hay que añadir que no se encontraron diferencias entre N+I y nivolumab para SG en la población global del estudio. Por tanto, ¿se puede asumir un beneficio asociado a N+I para SG en pacientes pre-tratados con CPNM escamoso avanzado, y con expresión de PD-L1 <1% y una TMB ≥ 10 mt/Mb? Para contestar a esta cuestión, se usará la herramienta desarrollada sobre aplicación de análisis por subgrupos.

3.3.4.2. Metodología

Se usó nuestra herramienta validada para evaluar la aplicabilidad del análisis por subgrupos con una metodología ordenada y sistemática (55). Esta herramienta está estructurada en dos partes: unas cuestiones preliminares para descartar la consideración de análisis por subgrupos sin condiciones mínimas relevantes, y un cuestionario. Las cuestiones preliminares tienen en cuenta el nivel de evidencia del

estudio que incluye el análisis por subgrupos, la relevancia de la variable abordada, la existencia de diferencia de efecto [$p(i) < 0.1$] (2) entre subgrupos para el factor seleccionado, y la secuencia temporal entre el factor y la intervención sanitaria. Una respuesta negativa ante alguna de las cuestiones preliminares conlleva al descarte de las siguientes preguntas, cuestionario y aplicabilidad del análisis por subgrupos. El cuestionario otorga una valoración y puntuación a cada uno de los criterios establecidos para interpretar el análisis por subgrupos: asociación estadística, que engloba la valoración de $p(i)$, pre-especificación de subgrupos, tamaño muestral, número de factores evaluados y resultado global del estudio; plausibilidad biológica y consistencia de los resultados. La puntuación es más elevada cuanto mayor sea la fiabilidad del criterio: “probable” (+3 puntos), “posible” (+2), “dudosa” (0) y “nula” (-3). La suma de las puntuaciones de los criterios se relaciona con una recomendación sobre la aplicabilidad de los resultados en la práctica clínica. Puntuaciones más altas de la herramienta se corresponden con una mayor aplicabilidad de los resultados de los subgrupos y, consecuentemente, con recomendaciones para que deban usarse en la toma de decisiones clínicas. La suma de las puntuaciones se asocian a las siguientes recomendaciones: 9-7 puntos (recomendación “probable”) se relaciona con aplicabilidad del análisis por subgrupos; 5-6 puntos (recomendación “posible”) con aplicabilidad cautelosa por limitada tolerancia, dificultad de uso o coste de otras intervenciones; 3-4 puntos (recomendación “dudosa”) se corresponde con evitar la aplicación salvo excepciones; y <3 puntos (recomendación “nula”) se asocia con evitar la aplicación de los resultados de subgrupos. Si la asociación estadística o la consistencia tuvieran una consideración de “nulas”, la aplicabilidad del análisis sería descartada directamente sin considerar el sumatorio de los criterios (recomendación “nula”).

Para calcular los valores de $p(i)$ entre subgrupos complementarios cuando estos datos no fueron aportados por el ensayo, se utilizaron calculadoras basadas en la escala del Ln de valores de riesgo relativo, odds ratio, hazard ratio e intervalos de confianza disponibles (12,33,68–72).

3.3.4.3. Resultados

El estudio abordado no proporcionó la estimación de $p(i)$. Se estimó el valor de $p(i)$ entre el subgrupo con expresión de PD-L1 <1% y TMB < 10 mt/MB, y el subgrupo con expresión de PD-L1 <1% y TMB \geq 10 mt/Mb. El valor de $p(i)$ fue 0,0043.

Posteriormente, se usó nuestra herramienta de interpretación de análisis por subgrupos sobre la afirmación de la existencia de una mejora de SG asociada al esquema N+I en pacientes pre-tratados con CPNM escamoso avanzado, y con expresión de PD-L1 <1% y una TMB \geq 10 mt/Mb. Se respondió positivamente a las cuestiones preliminares de la herramienta. Sin embargo, el análisis de la expresión de PD-L1 para el corte del 1% y diferentes niveles de TMB tiene un carácter *post hoc* y fue desarrollado en un escaso tamaño muestral. Ni siquiera se alcanzó una distribución de más de 50 pacientes por rama en los subgrupos. N+I aportó 9 pacientes al subgrupo con expresión de PD-L1 <1% y TMB <10 mt/Mb, mientras que el brazo de nivolumab en monoterapia contribuyó con 17 casos. El subgrupo con expresión de PD-L1 <1% y una TMB \geq 10 mt/Mb estuvo compuesto por 15 pacientes de cada brazo. A esto hay que añadir que el ensayo Lung-MAP S1400I no encontró beneficio de SG asociado a N+I en comparación con nivolumab en el global de pacientes con CPNM escamoso avanzado tratado previamente. Es decir, el resultado global del estudio fue negativo. Esto resta credibilidad a los resultados del análisis por subgrupos del ensayo clínico teniendo en cuenta su naturaleza exploratoria.

Teniendo todo esto en cuenta, el criterio de “asociación estadística” alcanzó una consideración de “nula”, por lo que la aplicabilidad del análisis se descartó directamente sin considerar el sumatorio de los criterios (recomendación “nula”). Por tanto, no sería necesario seguir evaluando el resto de criterios. Se decidió no continuar aplicando el resto de la herramienta y se descartó la aplicabilidad del análisis por subgrupos (**Tabla 14**).

Tabla 14. Resumen de evaluación sobre análisis por subgrupos con expresión de PD-L1 <1% y corte de TMB en 10 mt/Mb

Metodología	Criterios		Gettinger <i>et al.</i> (ensayo Lung-MAP S1400I), evaluación de SG†
Herramienta validada (Gil-Sierra <i>et al.</i>) (55)	Cuestiones preliminares	Nivel más alto de evidencia del estudio con análisis por subgrupos	Sí
		Relevancia clínica clara de variable considerada o variable primaria subrogada	Sí
		Existencia de diferencia en el efecto entre los subgrupos para el factor analizado [p(i) < 0.1]	Sí
		Factor determinante del análisis por subgrupos previo a la intervención sanitaria	Sí
		Cuestionario Asociación estadística (puntuación) Plausibilidad biológica (puntuación) Consistencia (puntuación) Recomendación	Nula (-3 puntos) No aplicado No aplicado No aplicado Nula: “Análisis por subgrupos no deben ser considerados”

†SG: supervivencia global

3.3.4.4. Discusión

Los resultados obtenidos al aplicar nuestra herramienta de interpretación de análisis por subgrupos indican que no se debería aceptar de forma concluyente un beneficio de N+I en SG para pacientes pre-tratados con CPNM escamoso avanzado, expresión de PD-L1 <1% y TMB ≥ 10 mt/Mb (55,104). Este es un claro ejemplo de que no siempre se puede asumir diferencias entre subgrupos complementarios simplemente porque $p(i)$ sea estimado en un valor menor a 0,1 o 0,05 (en el mejor de los casos). El análisis por subgrupos presenta una serie de limitaciones que deben tenerse en cuenta antes de realizar afirmaciones sobre la toma de decisiones clínica (2,4,5).

Una excesiva multiplicidad y desbalanceamiento de la redistribución de factores relacionados con el beneficio en el análisis por subgrupos aumenta la posibilidad de encontrar aparentes diferencias, que en realidad son inexistentes (55). Esto está relacionado con el elevado error α del análisis por subgrupos. Por otra parte, la pre-especificación del factor considerado en el análisis por subgrupos es importante para evitar que se desarrollen interpretaciones *post hoc* o *ad hoc* que favorezcan a la comercialización y la venta de determinados fármacos, cuando realmente no aportan beneficio relevante en un escenario clínico. Además, un reducido tamaño muestral minimiza el poder estadístico de un análisis por subgrupos y está relacionado con un elevado error β . La aplicación de resultados de subgrupos en la práctica clínica debería considerar subpoblaciones con, al menos, 50 pacientes de cada brazo. Este no es el caso del estudio Lung-MAP S1400I, donde los subgrupos complementarios con expresión de PD-L1 <1% y corte en 10 mt/Mb de TMB incluyen menos de 20 pacientes de cada brazo. Se trata de un tamaño muestral muy reducido. A todo esto, hay que sumar que el resultado global del estudio es negativo, es decir, no se observaron diferencias estadísticamente significativas entre la combinación N+I y nivolumab en monoterapia para pacientes pre-tratados con CPNM escamoso avanzado. Si ya el objetivo principal del ensayo no fue alcanzado, la consideración de un análisis exploratorio con resultados de subgrupos conlleva un grado de incertidumbre demasiado importante. Por ello, aún

no se debería de tener en cuenta los datos de los subgrupos según los biomarcadores PD-L1 y TMB a la hora de seleccionar el tratamiento para estos pacientes.

Si bien es cierto que se podría haber continuado la evaluación de los criterios de “plausibilidad biológica” y “consistencia”, también hay que señalar que sus resultados no tendrían ningún impacto en la recomendación sobre el modo de actuar frente al análisis por subgrupos. Siempre que la “asociación estadística” o la “consistencia” sean considerados “nulos”, se debería desechar la aplicación del análisis por subgrupos en la toma de decisiones clínicas. Por ello, debido a las limitaciones comentadas por los datos de los subgrupos de este caso concreto, se decidió no continuar usando la herramienta desarrollada.

Además de valorar la eficacia de los tratamientos, el peor perfil de seguridad del esquema N+I frente nivolumab también debe ser evaluado durante el posicionamiento terapéutico. Los eventos adversos grado 3-5, eventos relacionados con la inmunidad y discontinuaciones de tratamiento por escasa tolerancia se observaron más frecuentemente en la combinación de N+I. En este sentido, la mediana del tiempo de discontinuación del tratamiento N+I fue más prolongada que la de nivolumab en monoterapia. Un meta-análisis con más de 19000 pacientes mostró la mayor tasa de letalidad relacionada con la toxicidad en la combinación de anticuerpos PD-1/PD-L1 con agente inhibidor del punto de control inmunológico del antígeno 4 asociado al linfocito T citotóxico, entre otras inmunoterapias (113). La evaluación fármaco-económica también tiene un papel fundamental en la selección de medicamentos. Un estudio de coste-efectividad en CPNM avanzado determinó una pobre eficiencia para N+I en pacientes con expresión de PD-L1 menor al 1% en comparación con subgrupos de mayores porcentajes de expresión (114).

TMB es un biomarcador emergente interesante, aunque se necesitan ECA con resultados de subgrupos más consistentes para determinar en su influencia. Una mejora de SG asociada al uso del esquema N+I aún no debería predecirse en CPNM escamoso

avanzado con expresión de PD-L1 <1% y alta TMB. A esto hay que añadir que se deben reportar datos de tolerancia y eficiencia más favorables para la combinación N+I antes de ser considerada una alternativa terapéutica beneficiosa en este subgrupo de pacientes.

3.3.5. Remdesivir y reducción de mortalidad en pacientes con COVID-19.

Alegre-Del Rey EJ, Gil-Sierra MD, Alarcón de la Lastra-Romero C, Sánchez-Hidalgo M. Remdesivir and mortality reduction in COVID-19 patients: a systematized subgroup analysis of clinical trials. *Farm Hosp.* 2021;45(1):28-31. Doi: 10.7399/fh.11591.

3.3.5.1. Introducción

A finales del año 2020, se pre-publicaron los resultados del ensayo clínico SOLIDARITY de la OMS (105). En este estudio, remdesivir volvía a fallar al intentar mostrar beneficio en supervivencia a los 28 días en pacientes diagnosticados de COVID-19 grave. Estos resultados fueron valorados como variable principal y con mayor número de pacientes que en publicaciones previas, ya que esta falta de efecto en mortalidad había sido evaluada con menor potencia estadística en el ensayo ACTT-1 (93), ensayo SIMPLE (91) y Wang *et al* (31).

Los negativos efectos de la pandemia por COVID-19 favorecieron la realización de afirmaciones prematuras sobre resultados de nuevos fármacos, incluyendo datos sobre análisis por subgrupos. Los resultados por subgrupos del estudio ACTT-1 han sido usados para sugerir la hipótesis de una posible reducción de mortalidad en pacientes con COVID-19 grave que requieren oxígeno -no de alto flujo (NAF)- y que recibieron remdesivir (93). Podría parecer razonable que un antiviral tenga mayor efecto de control de la enfermedad en una fase temprana, antes de que una respuesta inmune descontrolada haga ya casi inútil una cierta supresión viral. La posibilidad de que remdesivir tuviera algún efecto positivo en el desarrollo de la COVID-19 parecía

evidente, especialmente en pacientes sin requerimiento absoluto de soporte ventilatorio. No obstante, hasta ese momento no se había observado que ese efecto se hubiera traducido en salvar la vida de pacientes, a pesar de que ya se disponía de datos de 7.600 casos aleatorizados en cuatro ensayos clínicos (31,91,93,105). No hay que olvidar que la reducción de la mortalidad es el objetivo principal de cualquier tratamiento contra la COVID-19, y que hasta ese momento sólo dexametasona lo había conseguido (86).

Por tanto, se habían desarrollado ECA sobre el uso de remdesvir en COVID-19 grave sin beneficio en el resultado global de mortalidad, y arrojando serias dudas sobre el beneficio del fármaco en el subgrupo de pacientes hospitalizados que requieren oxígeno sin alto flujo (31,91,93,105). Consecuentemente, era interesante evaluar los subgrupos del estudio SOLIDARITY para obtener respuestas, además de realizar un meta-análisis de subgrupos de los ECA publicados (105). Más si cabe, teniendo en cuenta su impacto sociosanitario y la rapidez de la divulgación de los datos –a pesar de ser previos a la publicación en la revista científica–. La versión pre-publicada del estudio SOLIDARITY mostraba un meta-análisis en el que se combinaron los resultados de los subgrupos de los cuatro ECA sobre la COVID-19 grave que había disponibles. Se subdividía a la población global de estos estudios en grupos de alto y bajo riesgo (que no reciben ventilación), incluyéndose a pacientes con oxígeno de alto flujo/NAF del estudio SOLIDARITY en este último grupo. La valoración de los datos de este meta-análisis arrojaría cierta información sobre el beneficio de mortalidad según el riesgo de los pacientes, contribuyendo a mejorar el conocimiento sobre el efecto de remdesvir en la mortalidad de pacientes que requieren oxígeno NAF según toda la evidencia científica disponible hasta ese momento.

Como ya hemos comentado en varias ocasiones, la aplicación del análisis de subgrupos conlleva asumir un incremento considerable de incertidumbre por la realización de determinaciones adicionales y nuevas redistribuciones de pacientes en las ramas (5). La interpretación del análisis por subgrupos es un asunto controvertido que requiere de una valoración sistemática y metodológica antes de la toma de decisiones clínicas (23).

Teniendo en cuenta todo lo anterior, el objetivo de este trabajo fue la evaluación metodológica del efecto de remdesivir en mortalidad a los 28 días presentado en el meta-análisis por subgrupos de los ECA disponibles en octubre de 2020, según el riesgo de los pacientes diagnosticados de COVID-19 grave.

3.3.5.2. Metodología

La interpretación del meta-análisis de subgrupos realizado en el estudio SOLIDARITY se llevó a cabo mediante nuestra herramienta validada que permite valorar la aplicabilidad del análisis por subgrupos (55). La metodología utilizada consta de dos partes: cuestiones preliminares que descartan la valoración de análisis por subgrupos sin condiciones mínimas relevantes, y un cuestionario. La respuesta negativa de alguna cuestión preliminar lleva a rechazar la aplicabilidad del análisis por subgrupos, sin necesidad de aplicar el cuestionario. Este cuestionario valora una serie de criterios para interpretar el análisis por subgrupos: asociación estadística, que incluye $p(i)$, pre-especificación de subgrupos, tamaño muestral, número de factores abordados y resultado global del estudio; plausibilidad biológica de las diferencias encontradas entre subgrupos; y consistencia entre resultados de estudios similares. Se asignó una puntuación correspondiente a cada valoración de los criterios establecidos en las cuestiones de asociación estadística, plausibilidad biológica y consistencia. Para ello, se usaron los siguientes rangos: “probable” (+3 puntos), “posible” (+2 puntos), “dudosa” (0 puntos) y “nula” (-3 puntos). La suma global se asoció con una recomendación específica sobre la aplicabilidad del análisis por subgrupos en la toma de decisiones clínicas. Esta recomendación podría seguirse siempre que la asociación estadística o consistencia no fueran “nulas”, lo que conllevaría el descarte directo de los resultados de los subgrupos. Una mayor puntuación global se corresponde con una mejor fiabilidad del análisis por subgrupos: “probable” (9-7 puntos) se asocia con la aplicabilidad del análisis por subgrupos hasta que se disponga de otro ECA confirmatorio; “posible” (5-6 puntos) con aplicabilidad cautelosa ante casos de escasa tolerancia, dificultad de uso o coste de alternativas terapéuticas; “dudosa” (3-4 puntos) se relaciona con el rechazo de

aplicabilidad salvo excepciones; y por último, “nula” (<3 puntos) con la no aplicación del análisis por subgrupos.

Si el artículo no aportó el valor de $p(i)$, se estimó mediante calculadoras de subgrupos basadas en la escala del Ln de valores de riesgo relativo, odds ratio, hazard ratio e intervalos de confianza (12,33,68–72).

3.3.5.3. Resultados

Las cuestiones preliminares de nuestra herramienta tuvieron respuestas positivas, por lo que se aplicó el cuestionario. El subgrupo de pacientes sin ventilación anexo al estudio SOLIDARITY presentó valores de razón de tasas de mortalidad (RdR) = 0.8 [95% IC, 0.63-1.01]. Por otra parte, el subgrupo de pacientes de alto riesgo obtuvo RdR = 1.16 [95% IC, 0.85-1.6]. Con respecto a la asociación estadística, aplicada al resultado global del meta-análisis en pacientes sin ventilación o sin oxígeno de alto flujo frente al resto, se calculó una $p(i) = 0,0650$ entre los subgrupos. Con este valor, la herramienta asigna una valoración “dudosa” a $p(i)$. El meta-análisis recrea el grupo de pacientes que requieren oxígeno NAF. Sin embargo, el estudio SOLIDARITY no presentaba pre-especificación para la evaluación de este subgrupo. Se realizó el análisis por subgrupos según el riesgo de los pacientes considerando el hallazgo de la aparente diferencia de mortalidad del ensayo ACTT-1 (93). Por tanto, se asignó la valoración de fiabilidad “nula” para la pre-especificación. El tamaño muestral es muy superior a 100 pacientes en ambos subgrupos, asociándose una valoración “probable” en este aspecto. Para valorar el número de factores analizados se consideraron los factores del estudio ACTT-1 (93), que es el “promotor” del análisis de subgrupos posterior del meta-análisis. Además, en el meta-análisis no se pueden valorar los factores pre-especificados ya que valora los establecidos en los ECA que incluye. El ensayo ACTT-1 llevó a cabo un análisis de 7 factores (<10), con una valoración de “probable” (93). La mortalidad no alcanzó significación estadística para la población global en el meta-análisis de los cuatro ensayos (RdR = 0.91; 95% IC, 0.79-1.05). Esto hace que estuviéramos en la posible situación de un resultado global negativo, que podría alcanzar un resultado positivo en un determinado subgrupo diferenciado. No obstante, este último escenario tampoco

llegó a verificarse. Ni siquiera en el subgrupo de pacientes sin ventilación/oxígeno de alto flujo del meta-análisis llega a alcanzarse significación estadística (RdR = 0.80; 95% IC, 0.63-1.01). Todo esto lleva a considerar que la asociación estadística reciba una valoración “nula”, pues ni se ha observado pre-especificación de subgrupos ni siquiera se cumple el presupuesto de este apartado (estudio negativo en población global, pero con diferencia estadísticamente significativa en uno de los subgrupos).

La plausibilidad biológica del análisis por subgrupos puede recibir una consideración de “probable”. Es posible que el antiviral tenga mayor efecto cuando aún no se ha desencadenado el proceso inflamatorio en su totalidad, que es el que produce el empeoramiento clínico en pacientes en estadios iniciales de la enfermedad. Estos pacientes están representados por el subgrupo de bajo riesgo del meta-análisis, que incluye a los pacientes que requieren oxígeno NAF. La consistencia recibió una valoración de “posible”. El análisis con interacción del ensayo ACTT-1 (93) no fue incompatible con los resultados mostrados en el estudio SOLIDARITY, los cuales no presentaron interacción, aunque apuntaron en la misma dirección. Los estudios de Wang *et al.* y SIMPLE no mostraron consistencia (31,91). Hay que tener en cuenta que su escasa potencia estadística hace que no puedan considerarse tampoco de forma concluyente.

Analizando todo lo anterior mediante la herramienta desarrollada (55), la recomendación de aplicación del análisis por subgrupos según el riesgo de los pacientes fue “nula”. Esto se debe al descarte directo causado por una asociación estadística de nula fiabilidad. La **Tabla 15** resume la interpretación del análisis por subgrupos según el estado del paciente aportado por el meta-análisis que incorpora resultados del estudio SOLIDARITY.

Tabla 15. Resumen de evaluación del análisis por subgrupos según riesgo descrito en el estudio SOLIDARITY

Metodología	Criterios		Ensayo SOLIDARITY, evaluación de mortalidad a los 28 días
Herramienta validada (Gil-Sierra <i>et al.</i>) (55)	Cuestiones preliminares	Nivel más alto de evidencia del estudio con análisis por subgrupos	Sí
		Relevancia clínica clara de variable considerada o variable primaria subrogada	Sí
		Existencia de diferencia en el efecto entre los subgrupos para el factor analizado [$p(i) < 0.1$]	Sí
		Factor determinante del análisis por subgrupos previo a la intervención sanitaria	Sí
		Cuestionario Asociación estadística (puntuación) Plausibilidad biológica (puntuación) Consistencia (puntuación) Recomendación de aplicación (sumatorio global)	Aplicado Nula (-3 puntos) Probable (+3 puntos) Posible (+2 puntos) Nula por asociación estadística nula (descarte directo): “Análisis por subgrupos no deben ser considerados”

3.3.5.4. Discusión

El resultado final de nuestra herramienta validada sobre aplicación de resultados de subgrupos fue “nulo” con respecto a la reducción de mortalidad asociada a remdesivir en pacientes con COVID-19 hospitalizados que reciben oxígeno NAF. Por tanto, podría afirmarse que esa aparente diferencia en mortalidad entre subgrupos pudiera haber sido compatible con el azar. No cabría aplicar –hasta ese momento– otro valor a la mortalidad que el global del meta-análisis asociado al estudio SOLIDARITY, es decir, $RdR = 0.91$ (95% IC, 0.79-1.05). No obstante, la nula fiabilidad de este análisis de subgrupos con información precaria no fue capaz de descartar de forma concluyente la hipótesis de un mayor beneficio en el subgrupo de pacientes que requieren oxígeno NAF.

La estimación de la magnitud del hipotético beneficio que podría ser obtenido en el subgrupo de bajo riesgo del meta-análisis adjunto al estudio SOLIDARITY (105) –el cual incluye el subgrupo de pacientes del estudio ACTT-1 que requieren oxígeno NAF para el que se ha sugerido una posible reducción de mortalidad (93)– se podría realizar mediante el número necesario de pacientes a tratar para reducir un evento adicional (NNT, del inglés *number needed to treat*) y la reducción relativa del riesgo (RRR). Si se llegase a confirmar un beneficio significativo, lo cual no se había establecido en el meta-análisis hasta ese momento, la mortalidad en el subgrupo de pacientes no intubados/con oxígeno NAF podría descender del 8.6% al 7.0%. Esto supondría una reducción absoluta del riesgo del 1.6%. De esta manera, se evitaría hipotéticamente una muerte adicional para cada 62 pacientes tratados con remdesivir (NNT). De forma global, y en términos relativos, evitaríamos una de cada cinco muertes en estos pacientes (1.6% sobre 8.6%, es decir, una RRR del 18.6%). La magnitud del beneficio obtenido en el subgrupo de bajo riesgo del estudio SOLIDARITY (105) apuntó a una dudosa mejoría en mortalidad en pacientes no intubados que, de existir, podría ser muy inferior a lo hallado aparentemente en el subgrupo de pacientes que requieren oxígeno NAF del estudio ACTT-1 (93).

Se encontró una limitación importante del meta-análisis del estudio SOLIDARITY (105): la diferente definición de los subgrupos en los cuatro ECA incluidos (31,91,93,105). No se pudo examinar el subgrupo de oxígeno NAF de forma aislada, que es el que ofreció un aparente beneficio relevante en el estudio ACTT-1 (93). Se tuvieron que limitar a agrupar pacientes sin ventilación y pacientes con ella. Los pacientes con oxígeno de alto flujo o ventilación no invasiva del ACTT-1 que no se beneficiaban fueron incorporados al grupo de pacientes con ventilación, y esto favoreció la aparición de diferencias entre los subgrupos de alto y bajo riesgo en el meta-análisis del ensayo SOLIDARITY(105). Por otra parte, el subgrupo que mezcla pacientes con oxígeno de alto flujo/NAF del estudio SOLIDARITY [RdR = 0.85, 95% IC 0.66-1.09] fue incluido en el grupo de los no ventilados.

Este trabajo ayudó a contextualizar los resultados observados en el ensayo ACTT-1 (93), en el que el subgrupo de pacientes con oxígeno NAF presentó un valor de $p(i) < 0.05$. Este hallazgo era biológicamente plausible, aunque el análisis no pareciera consistente con otros ECA previamente publicados (31,91). Lejos de caer en el típico error de la valoración de análisis por subgrupos sobre sacar conclusiones atendiendo a si los IC cruzan el valor neutro (10,12,21), se ha realizado una interpretación sistemática de los resultados más relevantes (meta-análisis) del efecto de remdesivir en pacientes con COVID-19 hospitalizados según el riesgo y estado clínico.

La situación de emergencia por la pandemia supuso un desafío sin precedentes para los profesionales sanitarios (115). La carencia de alternativas terapéuticas eficaces provocó que en determinados momentos pudiera ser aceptable el uso de remdesivir en pacientes con oxígeno NAF. Este fármaco fue empleado aun sin la certeza de estar salvando vidas, y con el riesgo de no estar haciendo más que adelantar la recuperación de los casos que se recuperarían de todas formas. La evidencia no nos permitía sugerir que remdesivir estuviera contribuyendo a reducir la mortalidad de la COVID-19, siendo éste nuestro principal objetivo. Esto hacía que fuera necesario investigar nuevos tratamientos, pautas o formas de uso que realmente estuvieran orientadas a salvar la vida de pacientes mediante un adecuado diseño de ECA. Para ello, era esencial seleccionar la mortalidad a 28 días como variable principal y no otras variables, como

una relativa recuperación. Esto no contribuyó a acelerar el conocimiento ni la disponibilidad de verdaderas soluciones.

Se ha llegado a la conclusión de que era probable que la reducción de mortalidad en pacientes con oxígeno NAF no pasara de ser un hallazgo casual. Aunque el beneficio fuera real, los resultados del meta-análisis estudiado señalan que hubiera sido más modesto que lo que aparentaba el hallazgo aislado del estudio ACTT-1. El análisis sistematizado de la variable de mortalidad descartó el uso de remdesivir en cualquiera de los subgrupos estudiados asumiendo un beneficio en supervivencia. Al abrirse prematuramente la puerta al uso de remdesivir en pacientes con oxígeno NAF por la excepcionalidad de la situación descrita, se podría haber exigido la realización de un ECA sobre mortalidad en este subgrupo para sustentar las decisiones clínicas con la evidencia imprescindible.

3.4. Estudio fármaco-económico sobre subgrupos de mepolizumab en asma eosinofílica refractaria grave

García-Mochón L, Gil-Sierra MD, Alegre-Del Rey EJ, Alarcón de la Lastra-Romero C, Sánchez-Hidalgo M. Economic evaluation and budgetary burden of mepolizumab in severe refractory eosinophilic asthma. *Farm Hosp.* 2019;43(6):187-93. Doi:

10.7399/fh.11221.

3.4. ESTUDIO FÁRMACO-ECONÓMICO SOBRE SUBGRUPOS DE MEPOLIZUMAB EN ASMA EOSINOFÍLICA REFRACTARIA GRAVE.

3.4.1. Introducción

Se ha estimado que el asma afecta aproximadamente al 4,9% de los adultos (116). En nuestro ámbito, la prevalencia del asma grave no controlada o refractaria al tratamiento con corticoides y agonistas β_2 de larga duración (LABA, del inglés *long-acting beta-adrenoceptor agonist*) supone aproximadamente el 3,9% de los asmáticos (117). Dentro de este grupo, aproximadamente el 25% de los casos padece asma eosinofílica. Esta patología se asocia con un inicio tardío, presencia de eosinófilos en biopsias bronquiales, y aparición de pólipos nasales, rinosinusitis e infecciones respiratorias (118,119).

Omalizumab es un anticuerpo indicado en asma alérgica grave no controlada que se autorizó en España en el año 2006 (120). La posología de omalizumab es variable. Los esquemas pueden oscilar desde un mínimo de 75 mg cada 4 semanas hasta un máximo de 600 mg cada 2 semanas (120). En el año 2015, se comercializó mepolizumab. Este anticuerpo monoclonal está indicado como tratamiento adicional del asma eosinofílica refractaria grave (121). Mepolizumab se une a la interleucina 5 e impide su interacción con la superficie de los eosinófilos. Esto provoca una reducción en la producción y supervivencia de los mismos. La dosis recomendada es de 100 mg cada 4 semanas. Existen estudios que evalúan la dosis de mepolizumab y la respuesta eosinofílica, demostrando una farmacodinámica similar entre la dosis de 100 mg y 75 mg (122).

Los ensayos clínicos pivotaes para la autorización de mepolizumab midieron su efecto sobre la frecuencia de exacerbaciones clínicamente relevantes como variable principal (30,37). Una exacerbación clínicamente relevante fue definida como aquel episodio agudo asmático que requiere uso de corticosteroides sistémicos durante al menos tres días y/u hospitalización y/o visita a urgencias, o corticosteroides sistémicos al doble de dosis de mantenimiento durante tres días o más en pacientes tratados con corticoides orales en mantenimiento (30,37). Mepolizumab ha demostrado reducir las

exacerbaciones y la dosis diaria de corticoides sistémicos orales (CSO) en asma eosinofílica grave no adecuadamente controlada con dosis elevadas de corticoides sistémicos inhalados (CSI) + LABA y/o CSO (tratamiento habitual).

Por otra parte, se asocia una mayor frecuencia de ataques de asma con el recuento de eosinófilos elevado (>300-400 células/ μ L) (123,124). Los análisis por subgrupos de los estudios pivotaes sugieren que el beneficio relativo es mayor en pacientes con mayores niveles de eosinófilos en sangre (30,37). El análisis por subgrupos está pre-especificado y presenta interacción estadística, observándose que la diferencia es consistente en los estudios (122,125). Además, existe una probable plausibilidad biológica. Un inhibidor de la población de eosinófilos podría ejercer mayor efecto cuanto mayor sea la contribución de la eosinofilia al proceso asmático.

Aproximadamente el 30% de los pacientes diagnosticados de asma eosinofílica presentan signos y síntomas compatibles con fenotipo de asma alérgica persistente mediada por inmunoglobulina E (IgE) (126), cumpliendo criterios para recibir omalizumab. No obstante, hasta el desarrollo de este estudio no se encontró evidencia para decantarse por un tratamiento u otro en esta subpoblación (125).

Considerando las diferencias de los subgrupos según el recuento de eosinófilos en el asma eosinofílica refractaria grave, la existencia de pacientes candidatos a mepolizumab u omalizumab, y el impacto económico del uso de mepolizumab, era importante realizar un estudio de evaluación económica e impacto presupuestario que facilitara la toma de decisiones clínicas eficientes. Hay que señalar que el impacto económico de las indicaciones de los fármacos en las poblaciones globales puede ser más manejable cuando se realizan valoraciones económicas complementarias según el análisis por subgrupos. En el momento de elaboración de este trabajo, otros medicamentos de mecanismo de acción similar a mepolizumab –reslizumab y benralizumab– se encontraban pendientes de financiación y precio en España (127,128). No existían comparaciones de estos fármacos con mepolizumab y resultaba complejo establecer

diferencias entre ellos. Por tanto, la inclusión de estos tratamientos en este estudio no pudo ser realizada.

A continuación, vamos a describir la realización de un análisis de coste/eficacia e impacto presupuestario (AIP) sobre la introducción de mepolizumab como tratamiento del asma eosinofílica refractaria grave mediada y no mediada por niveles elevados de IgE en pacientes no adecuadamente controlados con dosis elevadas de CSI+LABA y/o CSO en España, y el impacto de los subgrupos según el recuento de eosinófilos.

3.4.2. Métodos

Se llevó a cabo un análisis de coste eficacia y AIP a nivel nacional desde la perspectiva del Sistema Nacional de Salud (SNS). Se cuantificaron exclusivamente los costes directos (€). El AIP se desarrolló para un período de 3 años (2018-2020). Los análisis fueron realizados teniendo en cuenta las últimas guías de evaluación económica y AIP (129,130).

3.4.2.1. Población de estudio

El estudio incluyó a pacientes mayores de 12 años con asma refractaria grave. Para el AIP, se usaron estimaciones de población adulta asmática y prevalencia del asma grave refractaria en España (116,117). Tras esto, se calculó el porcentaje de pacientes con asma grave refractaria al tratamiento con diagnóstico de asma eosinofílica, y la población con asma mediada por niveles altos de IgE candidata a tratamiento con omalizumab. Se usaron datos del Instituto Nacional de Estadística (131). Además, se realizó un AIP por subgrupos de población según recuento plasmático de eosinófilos (**Tabla 16**).

Tabla 16. Valores de prevalencia, eficacia y costes utilizados en AIP.

	Valor medio	Valor mínimo – máximo	Referencia
Datos de prevalencia de asma (%)^a			
Población con asma	4,9	--	Estudio Europeo del Asma (116)
Población con asma grave refractaria al tratamiento	3,9	--	Quirce S <i>et al</i> (117)
Población candidata eosinofílica	25,0	--	IPT mepolizumab (125)
Subgrupo con < 150 eosinófilos/ μ L	23,1	--	Informe EMA (122)
Subgrupo con 150 a < 300 eosinófilos/ μ L	26,1	--	Informe EMA (122)
Subgrupo con 300 a < 500 eosinófilos/ μ L	20,5	--	Informe EMA (122)
Subgrupo con células \geq 500 eosinófilos/ μ L	30,4	--	Informe EMA (122)
Población con asma eosinofílica y mediada por IgE (candidatos a omalizumab)	30,0	--	Estudio OSO (126)
Tasas de exacerbaciones clínicas relevantes en grupo CSO			
Población total con asma eosinofílica grave no controlada			
Exacerbaciones clínicas significativas	1,91	--	Informe EMA (122)
Exacerbaciones que requieren hospitalización o visita a urgencias	0,26	--	Informe EMA (122)
Exacerbaciones que requieren hospitalización	0,14	--	Informe EMA (122)
Subgrupos según niveles de eosinófilos			
< 150 eosinófilos/ μ L	1,73	--	Informe EMA (122)
150 a < 300 eosinófilos/ μ L	1,14	--	Informe EMA (122)
300 a < 500 eosinófilos/ μ L	1,64	--	Informe EMA (122)
\geq 500 eosinófilos/ μ L	2,49	--	Informe EMA (122)
Riesgo relativo de MEPO vs CSO en exacerbaciones clínicas relevantes			
Población total con asma eosinofílica grave no controlada			
Exacerbaciones clínicas	0,51	0,42-0,62 ^b	Informe EMA (122)
Exacerbaciones que requieren hospitalización o visita a urgencias	0,53	0,33- 0,8 ^b	Informe EMA (122)
Exacerbaciones que requieren hospitalización	0,5	0,28-0,89 ^b	Informe EMA (122)
Subgrupos según niveles de eosinófilos			
< 150 eosinófilos/ μ L	0,67	0,46-0,98 ^b	Informe EMA (122)
150 a < 300 eosinófilos/ μ L	0,72	0,47-1,1 ^b	Informe EMA (122)
300 a < 500 eosinófilos/ μ L	0,62	0,41-0,93 ^b	Informe EMA (122)
\geq 500 eosinófilos/ μ L	0,27	0,19-0,37 ^b	Informe EMA (122)
Costes de medicamentos (€)			
Mepolizumab	14.118	--	BotPlus (132)
Omalizumab (precio notificado)	14.402	--	BotPlus (132)
Omalizumab (precio alternativo con descuento)	13.321	--	--
Otros costes relacionados (€)			
Coste de asistencia en urgencias de hospital	153,15	122,5-183,78 ^c	Estudio AsmaCost (133)
Coste de estancia hospitalaria (coste/día)	407,57	326-489 ^c	Estudio AsmaCost (133)
Coste estancias hospitalarias por exacerbación	3.845,43	2.028 – 7.860 ^d	Estudio AsmaCost (133)

CSO: corticoides sistémicos orales. EMA: Agencia Europea del Medicamento, del inglés *European Medicines Agency*. IPT: Informe de Posicionamiento Terapéutico. MEPO: mepolizumab. ^aLos cálculos se realizaron para la población residente en España. Las proyecciones se calcularon en octubre de 2014, tomando como referencia la población provisional a 1 de enero de 2014 (última cifra de población disponible en ese momento). ^bValores mínimo y máximos según 95% IC de estudios. ^cValores mínimos y máximos asumiendo una variación del 20% con respecto a la media. ^dCoste medio correspondiente a una estancia media de 9 días. Los valores mínimo y máximo se estimaron asumiendo 5 y 12 días de ingreso, respectivamente.

3.4.2.2. Alternativas terapéuticas evaluadas

El análisis de coste/eficacia y AIP sobre el uso de mepolizumab se ha realizado mediante dos análisis distintos. En el análisis 1, se comparó la asociación de dosis elevadas de CSI+LABA y/o CSO junto a mepolizumab frente a la misma medicación asociada a omalizumab en asma eosinofílica alérgica y mediada por IgE. En el análisis 2, se evaluó el uso de dosis elevadas de CSI+LABA y/o CSO junto a mepolizumab frente a dosis elevadas de CSI+LABA y/o CSO en asma refractaria grave no mediada por IgE. Este segundo análisis no incluye el 30% de pacientes con asma eosinofílica (tratados con omalizumab).

La posología seleccionada de mepolizumab es de 100 mg cada 4 semanas (122). Omalizumab se dosifica en función del peso corporal y niveles de IgE. La dosis puede oscilar desde 75 mg cada 4 semanas hasta 600 mg cada 2 semanas (120). Para el análisis de costes, se utilizó un promedio de estos valores (**Tabla 16**).

3.4.2.3. Medida de los resultados en salud

La eficacia de los tratamientos fue obtenida del IPT de mepolizumab (125) y del informe de mepolizumab de la Agencia Europea del Medicamento (122). Fueron estimadas las exacerbaciones clínicas, aquellas que requieren hospitalización o visitas a urgencias y las exacerbaciones clínicas relevantes por subgrupos de población según el recuento plasmático de eosinófilos (**Tabla 16**). Se evaluó la eficacia de los fármacos mediante la reducción en la media de exacerbaciones anuales clínicamente relevantes. Un estudio de minimización de costes exige evidencia de equivalencia clínica de los fármacos valorados. En el momento del estudio, se carecía de evidencia clínica comparativa de calidad entre mepolizumab y omalizumab que demostrara equivalencia clínica o diferencia entre ambos tratamientos. Por tanto, no pudo realizarse un estudio de minimización de costes. En su lugar, se desarrolló un estudio de comparación de costes en el análisis 1.

3.4.2.4. Estimación de costes

Fueron incluidos el coste de la medicación (omalizumab y mepolizumab), exacerbación clínica relevante, visita a urgencias y hospitalización a causa de exacerbación asmática. Los fármacos fueron valorados mediante el precio de venta del laboratorio según el Catálogo de Especialidades Farmacéuticas del Consejo General de Colegios Oficiales de Farmacéuticos (132). Para omalizumab, se seleccionó su precio notificado y precio alternativo según práctica clínica habitual con hipotético descuento del 7,3%. El precio notificado y alternativo coincidieron en el caso de mepolizumab (**Tabla 16 y Tabla 17**). El estudio AsmaCost fue usado para extraer el coste asociado a una exacerbación clínicamente relevante que requiere de hospitalización y/o asistencia en urgencias, y el coste de hospitalización –asumiendo una estancia media de nueve días– (133). El análisis incluye costes directos de medicamentos y de asistencia en urgencias y hospitalización, debido al impacto en la definición de exacerbación clínicamente relevante. El estudio no incluye costes relacionados con la administración hospitalaria de fármacos.

El presente trabajo evaluó el coste-incremental y coste del tratamiento en el AIP del análisis 1, y el coste por exacerbación evitada y coste del tratamiento de la población a estudio en el análisis 2.

3.4.2.5. Análisis de escenarios e incertidumbre

En el análisis 1, se desarrollaron escenarios con diferentes porcentajes de introducción de mepolizumab en el mercado sustituyendo a omalizumab (50, 70 y 100%), y considerando diferentes precios de omalizumab (precio notificado y precio alternativo). En el análisis 2, se realizaron análisis de sensibilidad para valorar la incertidumbre en torno a valores mínimos y máximos del 95% IC, de riesgos relativos (RR) de las variables (exacerbación clínica relevante, hospitalización y asistencia a urgencias), así como de costes asociados a hospitalización y asistencia a urgencias (**Tabla 16**).

3.4.3. Resultados

La población del estudio se representa en la **Tabla 16**.

3.4.3.1. Análisis 1. Población con asma eosinofílica alérgica y mediada por IgE

Se calculó un coste-incremental medio de mepolizumab con respecto a omalizumab (precio alternativo) de 797 €/por paciente y año, aunque esto también depende de cada paciente y posología de omalizumab. En la **Tabla 17** se pueden consultar los datos del AIP según penetración en el mercado, precio notificado o alternativo y año para pacientes con asma eosinofílica mediada por IgE. Teniendo en cuenta el precio alternativo con descuento de omalizumab, el escenario de incluir mepolizumab en pacientes con asma eosinofílica alérgica y mediada por IgE hubiera supuesto un aumento aproximado del gasto público entre 2,3 y 4,6 millones €, según año y grado de penetración en el mercado de mepolizumab. El impacto presupuestario en tres años hubiera ascendido a un incremento aproximado de 10,3 millones € con penetración gradual en el mercado, o de 14 millones € en un escenario de completa sustitución de omalizumab por mepolizumab. Teniendo en cuenta el precio notificado de omalizumab (mayor al alternativo), la introducción gradual de mepolizumab en el SNS se hubiera traducido en un ahorro aproximado de 3,6 millones € en tres años. La sustitución completa de omalizumab por mepolizumab se hubiera traducido en una reducción del gasto público de unos 5 millones €.

3.4.3.2. Análisis 2. Población con asma refractaria grave no mediada por IgE

La **Tabla 18** detalla los datos de coste por exacerbación evitada en la población con asma refractaria grave eosinofílica no mediada por IgE, la cual representaba el 70% de la población susceptible de tratamiento y cuya alternativa terapéutica era CSI+LABA y/o CSO. El coste por exacerbación evitada al añadir mepolizumab fue 15.085 €. Los datos por subgrupos de pacientes según el nivel plasmático de eosinófilos mostraron que el coste fue de 7.767 € por exacerbación evitada para el grupo de mayor afectación clínica basal y mayor eficacia relativa (pacientes con ≥ 500 eosinófilos/ μL).

El análisis de sensibilidad permite observar que el RR es una variable muy sensible a los resultados por subgrupos. Al tomar valores máximos del RR, en subgrupos con <500 eosinófilos/ μ L se hubieran alcanzado costes por exacerbación evitada mayores de 100.000 € adicionales con respecto a la población general con asma eosinofílica no controlada. Por contra, en un escenario de valores mínimos de RR para el subgrupo de pacientes de 300 a <500 eosinófilos/ μ L, mepolizumab hubiera tenido un coste por exacerbación evitada de 14.590,7 €.

Tabla 17. Resultados del AIP en pacientes con asma eosinofílica alérgica y mediada por IgE (años 2018-2020).

Año	Población (candidatos)	Situación actual		Situación nueva	
		Coste OMA (€)	% sustitución de OMA por MEPO	Coste de MEPO y OMA (€)	AIP (coste de MEPO – coste de OMA) (€)
Coste de OMA a precio notificado. Penetración progresiva de MEPO en el mercado					
2018	5.858	84.358.056	50	83.527.657	-830.399,25
2019	5.868	84.506.625	70	83.342.019	-1.162.558,96
2020	5.880	84.677.546	100	83.010.457	-1.660.798,51
2018-2020		253.542.227	--	249.880.133	-3.653.757
Coste de OMA a precio notificado. Penetración de MEPO 100% desde el inicio					
2018	5.858	84.358.056	100	82.697.257	-1.660.798,51
2019	5.868	84.506.625	100	82.842.902	-1.663.723,47
2020	5.880	84.677.546	100	83.010.457	-1.667.088,46
2018-2020		253.542.227	--	248.550.616	-4.991.610
Coste de OMA a precio alternativo. Penetración progresiva de MEPO en el mercado					
2018	5.858	78.030.116	50	80.363.687	2.333.570,40
2019	5.868	78.167.542	70	81.440.294	3.272.752,32
2020	5.880	78.325.641	100	83.010.457	4.684.816,69
2018-2020		234.523.298	--	249.880.133	10.291.140
Coste de OMA a precio alternativo. Penetración de MEPO 100% desde el inicio					
2018	5.858	78.030.116	100	82.697.257	4.667.140,79
2019	5.868	78.167.542	100	82.842.902	4.675.360,46
2020	5.880	78.325.641	100	83.010.457	4.684.816,69
2018-2020		234.523.298	--	248.550.616	14.027.318

AIP: análisis de impacto presupuestario. MEPO: mepolizumab. OMA: omalizumab.

Tabla 18. Coste por exacerbación evitada en población con asma refractaria grave eosinofílica (pacientes con asma no mediada por IgE) en 2018.

	Análisis basal (valor medio de RR)			Análisis de sensibilidad (mejor escenario, valores mínimos de RR)		Análisis de sensibilidad (peor escenario, valores máximos de RR)	
	Coste MEPO (€)	Exacerbaciones evitadas	Coste / exacerbación evitada ^a (€)	Exacerbaciones evitadas	Coste / exacerbación evitada (€)	Exacerbaciones evitadas	Coste / exacerbación evitada (€)
Población con asma eosinofílica grave no controlada	14.118	0,936	15.085	1,10	12.744,2	0,726	19.451,6
Subgrupo: < 150 eosinófilos/ μ L	14.118	0,570	24.729,4	0,934	15.112,4	0,035	408.034,7
Subgrupo: 150 a < 300 eosinófilos/ μ L	14.118	0,320	44.229,3	0,604	23.366,4	-0,114	Dominado
Subgrupo: 300 a < 500 eosinófilos/ μ L	14.118	0,620	22.654	0,986	14.590,7	0,115	122.979,1
Subgrupo: \geq 500 eosinófilos/ μ L	14.118	1,820	7.767	2,017	6.999,9	1,569	8.999,8

MEPO: mepolizumab. RR: riesgo relativo. ^aMEPO + tratamiento habitual *versus* tratamiento habitual. Tratamiento habitual: corticoides sistémicos inhalados + agonista β 2 de larga duración + corticoides sistémicos orales.

La **Tabla 19** muestra los datos del AIP para población con asma grave eosinofílica no mediada por IgE y por subgrupos según niveles plasmáticos de eosinófilos en 2018-2020. El impacto presupuestario anual de la población con asma eosinofílica grave no mediada por IgE hubiera alcanzado aproximadamente los 189 millones € (568,1 millones € en tres años). Si sumamos estos costes al resultado de AIP en tres años de pacientes con asma eosinofílica mediada por IgE (30% del global de pacientes con asma eosinofílica), y asumiendo la penetración progresiva de mepolizumab en el mercado (aproximadamente 10,3 millones € según la **Tabla 17**), se hubiera obtenido un AIP total aproximado de 578,4 millones €.

El AIP del período 2018-2020 para la población con asma eosinofílica grave no mediada por IgE (**Tabla 19**) señaló unas hipotéticas estimaciones de 57,6 millones € anuales en el subgrupo con recuento de ≥ 500 eosinófilos/ μL (aproximadamente 173 millones € en tres años). Si sumáramos el resultado del AIP con introducción progresiva de mepolizumab en tres años en los pacientes con asma eosinofílica mediada por IgE al AIP del uso de mepolizumab en la población con asma eosinofílica no mediado por IgE con ≥ 500 eosinófilos/ μL , el AIP total en 3 años hubiera ascendido de 183,3 millones €.

La **Tabla 19** detalla un análisis de sensibilidad del impacto presupuestario del uso de mepolizumab para el subgrupo de pacientes con ≥ 500 eosinófilos/ μL . De esta manera, se puede ilustrar las variaciones en el AIP que se hubieran podido dar en el mejor y peor escenario de mepolizumab en este subgrupo. Se tuvieron en cuenta variaciones en los costes de urgencias y hospitalización, y RR de exacerbaciones clínicas relevantes. Se observa que el AIP en 3 años en esta subpoblación hubiera oscilado entre 167 y 173,5 millones €.

Tabla 19. AIP de mepolizumab en población con asma eosinofílica grave no mediada por IgE y según subgrupos durante 2018-2020 (análisis de sensibilidad en subgrupo \geq 500 eosinófilos/ μ L).

Año	Población	Coste MEPO (€)	Ahorro exacerbaciones evitadas (€)	AIP (€)
Población con asma eosinofílica grave no mediada por IgE no controlada				
2018	13668	192.960.266,98	3.934.849,62	189.025.417,36
2019	13692	193.300.104,59	3.941.779,60	189.358.325
2020	13719	193.691.066,81	3.949.752,11	189.741.314,70
	2018-2020	579.951.438,38	11.826.381,33	568.125.057,06
Subgrupo: < 150 eosinófilos/μL				
2018	3154	44.529.292	518.401,30	44.010.891,08
2019	3160	44.607.716	519.314,30	44.088.402,14
2020	3166	44.697.938	520.364,65	44.177.573,85
	2018-2020	133.834.946	1.558.080,25	132.276.867,07
Subgrupo: 150 a < 300 eosinófilos/μL				
2018	3564	50.321.721	980.748,43	49.340.972,23
2019	3571	50.410.346	982.475,70	49.427.870,53
2020	3578	50.512.304	984.462,82	49.527.841,66
	2018-2020	151.244.371	2.947.686,95	148.296.684,42
Subgrupo: 300 a < 500 eosinófilos/μL				
2018	2795	39.460.917,64	992.054,77	38.468.862,87
2019	2800	39.530.415,39	993.801,96	38.536.613,43
2020	2806	39.610.368,26	995.811,99	38.614.556,27
	2018-2020	118.601.701,29	2.981.668,72	115.620.032,57
Subgrupo: \geq 500 eosinófilos/μL				
2018	4154	58.648.336,31	1.100.769,67	57.547.566,63
2019	4161	58.751.626,54	1.102.708,32	57.648.918,21
2020	4170	58.870.455,58	1.104.938,62	57.765.516,95
	2018-2020	176.270.418,43	3.308.416,61	172.962.001,79
Subgrupo: \geq 500 eosinófilos/μL (valor máximo de RR y valores mínimos de costes hospitalarios y de asistencia en urgencias)				
2018	4154	58.648.336,31	144.528,13	57.740.906,78
2019	4161	58.751.626,54	144.782,67	57.842.598,87
2020	4170	58.870.455,58	145.075,50	57.959.589,34
	2018-2020	176.270.418,43	434.386,30	173.543.094,99
Subgrupo: \geq 500 eosinófilos/μL (valor mínimo de RR y valores máximos de costes hospitalarios y de asistencia en urgencias)				
2018	4154	58.648.336,31	3.092.238,16	55.556.098,15
2019	4161	58.751.626,54	3.097.684,14	55.653.942,39
2020	4170	58.870.455,58	3.103.949,41	55.766.506,17
	2018-2020	176.270.418,43	9.293.871,71	166.976.546,71

AIP: análisis de impacto presupuestario. MEPO: mepolizumab. RR: riesgo relativo.

3.4.4. Discusión

La aparición de medicamentos de elevado impacto económico hace que sean necesarios estudios económicos que faciliten la optimización de los recursos de los sistemas sanitarios (134). Nuestra evaluación económica comparó dos alternativas terapéuticas en un grupo de pacientes con asma eosinofílica que presentan signos compatibles con fenotipo de asma alérgica persistente mediada por IgE. El diseño de este análisis económico puede ayudar en la toma de decisiones clínicas para favorecer la eficiencia mediante competencia de precios.

El resultado en salud valorado fue el número de exacerbaciones clínicamente relevantes evitadas al tratar a los pacientes con mepolizumab. La variable seleccionada es adecuada para guiar la toma de decisiones clínicas. En otros estudios se consideró la disminución de ingresos en el hospital, visitas a urgencias o a facultativos de atención primaria (135–137). Por otra parte, las comparaciones realizadas con respecto a otras alternativas terapéuticas (omalizumab y dosis elevadas de CSI+LABA y/o CSO) mejoran la validez de este estudio.

Nuestro trabajo presentaba como limitación la falta de evidencia comparativa de calidad entre mepolizumab y omalizumab en pacientes con asma eosinofílica mediada por IgE. Esta población era candidata a ambos tratamientos y se presentaba como la intersección específica de los dos conjuntos. Hasta el momento de desarrollo de este trabajo, se habían publicado dos estudios (138,139) –uno de ellos financiado por el laboratorio GlaxoSmithKline– que realizaban una comparación indirecta de mepolizumab frente a omalizumab en asma eosinofílica con signos y síntomas compatibles con fenotipo de asma alérgica persistente. Ambos describían una ausencia de diferencia en eficacia entre mepolizumab y omalizumab, resaltando la imposibilidad de hacer recomendaciones de un uso prioritario de un fármaco frente a otro. Esto fue debido a la alta heterogeneidad entre ensayos clínicos y los diferentes criterios de selección del uso de los medicamentos. La aplicación de una comparación indirecta fiable no fue posible, ya que mepolizumab se evaluó en asma mediada por componente eosinofílico con

independencia de los niveles de IgE, y omalizumab se valoró en asma mediada por IgE con independencia del componente eosinofílico, usándose en pacientes con IgE elevada no respondedores a otros tratamientos. Estas limitaciones fueron reseñadas en informes de evaluación de mepolizumab de países como Canadá (140) y Reino Unido (141). Teniendo en cuenta todo lo anterior, no se realizó una minimización de costes. En su lugar, sería razonable seleccionar los fármacos mediante una comparación de costes, excepto en aquellos pacientes que por alguna razón clínica justificada se prefiera uno o evitar otro.

Durante el período de tiempo en que se desarrolló este estudio, se aprobaron otros dos anticuerpos con mecanismo de acción similar a mepolizumab (127, 128). Como aún no estaban comercializados en España, no pudieron ser objeto del presente análisis. Una vez comercializados, lo más lógico es evaluar si las posibles comparaciones indirectas detectan diferencias clínicamente relevantes, teniendo en cuenta el nivel plasmático de eosinófilos. Su introducción en terapéutica favorece la competencia y podría reducir el impacto presupuestario de estos medicamentos. Su no inclusión en este estudio fue una limitación que debe completarse en estudios posteriores, una vez que se ha introducido mepolizumab en el mercado. La ulterior comparación entre esos fármacos sería procedente, pero también compleja. Estos anticuerpos no se han comparado directamente, sus estudios tienen criterios de inclusión dispares y se definió de forma heterogénea a los subgrupos según el recuento sanguíneo de eosinófilos.

Estudios previos demostraron que pacientes con un recuento elevado de eosinófilos plasmáticos obtuvieron mayor beneficio en comparación con pacientes con niveles bajos (30,37). Este análisis económico encontró un coste eficacia-incremental más favorable en los pacientes con ≥ 500 eosinófilos/ μL con respecto a los que tenían un menor recuento. Se debe destacar que el análisis por subgrupos de los ensayos pivotaes cumple los criterios de interacción, pre-especificación, plausibilidad biológica y consistencia en los diferentes estudios (122,125). Las evaluaciones económicas de mepolizumab publicadas en población con asma eosinofílica refractaria, sin considerar

los análisis de subgrupos, concluyeron que este fármaco no es costo efectivo. De esta manera, se instó a realizar descuentos de precios del 60-70% para que su financiación por los sistemas sanitarios fuera recomendable (142,143). Bermejo *et al* detallaron el proceso de evaluación de mepolizumab por el National Institute for Health and Care Excellence (141). En su estudio de evaluación económica definieron la población objetivo según el grado de severidad del asma y con niveles ≥ 300 eosinófilos/ μL . En este caso, también se mostraron resultados no coste-efectivos para este subgrupo de pacientes. Se recomendó su utilización sólo cuando se procediera al descuento de precios acordado y confidencial que hiciera a este medicamento coste-efectivo para esta subpoblación.

Hasta el desarrollo de este trabajo se carecía de evidencia comparativa de calidad entre mepolizumab y omalizumab en asma con componente eosinofílico y mediación por IgE. Tampoco existían otras evaluaciones económicas que compararan estos dos medicamentos. Por este motivo, se realizó una comparación de costes en estos pacientes. Desde la perspectiva del SNS y teniendo en cuenta el elevado impacto económico de mepolizumab, se concluyó que sería razonable utilizar el fármaco de menor coste y promover una competencia de precios. Esta estrategia no excluiría la excepcionalidad de preferencia de un tratamiento concreto para algún paciente cuando estuviera debidamente justificada. Tras este análisis fármaco-económico, priorizar el uso de mepolizumab en pacientes con asma eosinofílica refractaria grave no mediada por IgE y niveles plasmáticos elevados de eosinófilos (≥ 500 células/ μL), tal y como se indicó en el IPT de mepolizumab de la Agencia Española del Medicamento, permitiría mejorar la eficiencia y disminuir el impacto presupuestario (125). Este trabajo demostró la influencia que puede tener un adecuado análisis fármaco-económico sobre un análisis por subgrupos en la sostenibilidad del sistema sanitario.

4. DISCUSIÓN GENERAL

Gil-Sierra MD, Briceño-Casado MDP, Sánchez-Hidalgo M. Clinical and pharmacoeconomic impact of subgroup analysis in onco-hematological patients. Support Care Cancer. 2022;30(5):3761-72. Doi: 10.1007/s00520-022-06823-y

4. DISCUSIÓN GENERAL

Esta Tesis Doctoral pone de manifiesto el considerable impacto clínico y fármaco-económico que pueden tener los análisis por subgrupos sobre los pacientes y los sistemas de salud. Se ha desarrollado una herramienta que permite la evaluación sistemática y ordenada de los datos de los subgrupos teniendo en cuenta criterios objetivos. Posteriormente, se ha aplicado a numerosas revisiones sistemáticas y publicaciones concretas en los que se realizaban interpretaciones inadecuadas sobre los análisis por subgrupos. Afirmaciones erróneas sobre los resultados de determinadas subpoblaciones podrían afectar a la eficacia, seguridad, adecuación y posicionamiento terapéutico de los tratamientos, además de que podría influir en la selección de datos para la realización de estudios posteriores. Por último, se ha llevado a cabo un análisis fármaco-económico sobre el uso de mepolizumab asma eosinofílica grave no mediada por IgE según el recuento de eosinófilos, de manera complementaria a un análisis de coste/eficacia y AIP sobre la introducción de mepolizumab como tratamiento del asma eosinofílica refractaria grave mediada y no mediada por niveles elevados de IgE. De esta manera, se ha observado la mejoría de eficiencia que se puede obtener mediante la aplicación adecuada de un estudio fármaco-económico sobre un análisis por subgrupos.

La efectividad de los tratamientos obtenida por los pacientes de determinados escenarios clínicos puede verse afectada –o al menos su percepción– al interpretar erróneamente los datos de subgrupos. Uno de los errores más relevantes es la falta de consideración del valor de $p(i)$ entre subgrupos complementarios. Lamentablemente, la mayoría de los ECA ni siquiera mencionan $p(i)$. Un ejemplo es el estudio KEYNOTE-189 (101). Los autores de este estudio sugirieron que los pacientes con CPNM no tratado previamente y una expresión de PD-L1 igual o superior al 50% obtenían un mayor beneficio de pembrolizumab más quimioterapia. Sin embargo, estimaciones posteriores de $p(i)$ para las variables de SG y SLP indicaron que no existían diferencias entre los pacientes con expresión de PD-L1 igual o mayor del 50% y los que mostraban expresión menor de estas cifras (110). La **Tabla 20** resume los datos por subgrupos según expresión PD-L1 del estudio KEYNOTE-189.

Tabla 20. Datos del análisis por subgrupos del estudio KEYNOTE-189.

Estudio	Población incluida en análisis por subgrupos	Grupo intervención	Grupo comparador	Variable	Subgrupo de población	Nº de eventos	Nº de pacientes	Riesgo relativo, hazard ratio (95% IC)	Interacción estadística entre subgrupos [p(i)]	Interpretación incorrecta	Interpretación adecuada	Relevancia del error de interpretación
Gandhi <i>et al</i> (ensayo KEYNOTE-189) (101)	616	Pembrolizumab 200 mg en combinación con quimioterapia basada en platino más pemetrexed, cada 3 semanas	Quimioterapia basada en platino más pemetrexed, cada 3 semanas	Supervivencia global	Expresión de PD-L1 del 1-49%	65	186	0,55 (0,34–0,90)	0,44	El mayor beneficio relativo fue demostrado en el subgrupo de pacientes con expresión PD-L1 ≥50%	No se observó heterogeneidad entre el subgrupo de pacientes con expresión de PD-L1 1-49% y el subgrupo con PD-L1 ≥50%. No se demostraron diferencias de eficacia entre estos subgrupos	Posible aumento de uso de pembrolizumab más quimioterapia en subgrupo con PD-L1 ≥50% sin base estadística científica sólida. No se reservaría quimioterapia basada en platino para líneas de tratamiento posteriores
					Expresión de PD-L1 ≥50%	70	202	0,42 (0,26–0,68)				
				Supervivencia libre de progresión	Expresión de PD-L1 del 1-49%	114	186	0,55 (0,37–0,81)	0,12			
					Expresión de PD-L1 ≥50%	186	202	0,36 (0,25–0,52)				

P(i) fue estimado usando calculadoras e información de bibliografía publicada previamente (12,33,68–72).

Otro caso similar se pudo observar en el ensayo MONARCH 2 (102). Este estudio evaluó el uso de abemaciclib asociado a fulvestrant en pacientes con cáncer de mama avanzado que presentó receptores hormonales positivos y ERBB-2 negativo, tras progresión a terapia endocrina. Los autores afirmaron que este esquema reportó mayores beneficios en los subgrupos de pacientes con metástasis viscerales y resistencia primaria a terapia endocrina. Estas interpretaciones se llevaron a cabo usando diferencias numéricas y gráficas aparentes en SG. No obstante, el valor de $p(i)$ no sugirió heterogeneidad entre estos subgrupos con respecto a sus correspondientes complementarios (144). La falta de datos esenciales de eficacia –como el citado $p(i)$ – sobre los análisis por subgrupos puede inducir a interpretaciones ambigüas. Esto, a su vez, puede desembocar en la divulgación de conclusiones inadecuadas sobre determinados fármacos. Todo ello puede favorecer a una mayor prescripción facultativa de un medicamento en un determinado contexto clínico sin una base estadística-científica sólida. Además, se pueden liderar falsas expectativas en los médicos y pacientes. La toma de decisiones clínicas sobre intervenciones sanitarias sin una mayor eficacia demostrada –o al menos, dudosa– en una subpoblación requiere de una evaluación adicional de seguridad, conveniencia y eficiencia.

El ensayo MAIA demostró un aumento de SLP por el uso de daratumumab asociado a lenalidomida y dexametasona en comparación a lenalidomida y dexametasona en pacientes con MM no candidatos a trasplante que no habían recibido tratamiento previo (45). Los autores de este estudio afirmaron que no se observó superioridad en los pacientes con insuficiencia hepática basándose en los datos de subgrupos. En esta publicación, la heterogeneidad observada fue baja para SLP entre el subgrupo de pacientes con función hepática normal y el subgrupo con disfunción hepática. Esta aparente diferencia entre subgrupos según la función hepática no se encontró ni para SLP ni SG en el ensayo ALCYONE, el cual comparó daratumumab con bortezomib, melfalán y prednisona frente a bortezomib, melfalán y prednisona en primera línea de tratamiento de MM sin opción a trasplante (7,52). La ausencia de diferencias consistentes entre los subgrupos según función hepática en estudios similares (ensayos ALCYONE y MAIA) sugiere que la baja heterogeneidad observada en el estudio MAIA puede deberse al azar o desbalanceamiento de factores relacionados con el beneficio.

Considerando las ya mencionadas limitaciones del análisis por subgrupos, y el beneficio en SLP y SG de los tratamientos basados en daratumumab, no sería recomendable descartar estas alternativas terapéuticas en pacientes con función hepática alterada.

Las interpretaciones inadecuadas sobre datos de subgrupos no sólo ocurren en el ámbito de la onco-hematología, sino que pueden producirse en cualquier tipo de escenario clínico como, por ejemplo, el de las enfermedades infecciosas. Tal y como mencionamos anteriormente, Wang *et al* fueron los primeros autores en sugerir un beneficio del uso temprano de remdesivir en pacientes con COVID-19 grave (31). Este análisis por subgrupos fue llevado a cabo usando un corte a los 10 días desde el inicio de los síntomas y no se detalló el valor de $p(i)$ para ninguna de las variables valoradas. Análisis posteriores encontraron diferencias de eficacia entre subgrupos según los días transcurridos desde el inicio de los síntomas (93), aunque el corte se hizo en el séptimo día y sólo se observó heterogeneidad para variables de mejoría clínica, pero no para variables finales como SG. Por tanto, aplicar en la práctica clínica diaria las afirmaciones sobre un beneficio claro de remdesivir en COVID-19 grave antes de los 10 días hubiera sido un error. Se estaría sobre-estimando la eficacia de un tratamiento, ya que el mayor beneficio se obtiene antes de los 7 días desde el inicio de los síntomas y no se reportaron mejoras de SG, sino sólo una mejoría clínica más temprana.

Asimismo, también se sugirió –a partir de los resultados por subgrupos del estudio ACTT-1 (93)– una posible reducción de mortalidad en pacientes con COVID-19 grave que requieren oxígeno NAF y que recibieron remdesivir. La valoración de publicaciones de mayor entidad –estudio SOLIDARITY (105)– con nuestra herramienta sobre subgrupos indicó que, hasta ese momento, aún no existía evidencia científica suficiente para relacionar este beneficio con el uso de remdesivir. De hecho, aunque se asumiera esta aparente reducción de mortalidad, la magnitud del beneficio en el subgrupo de bajo riesgo del estudio SOLIDARITY (105) podría ser muy inferior a lo observado en el subgrupo de pacientes con oxígeno NAF del estudio ACTT-1 (93). Como podemos comprobar, la crisis sanitaria causada por la COVID-19 ha sido un caldo de cultivo perfecto para que interpretaciones inadecuadas sobre análisis por subgrupos generen falsas percepciones sobre la eficacia de tratamientos. Esto ha podido desembocar en un

uso excesivo de terapias de efecto dudoso –o directamente ineficaces– con elevado impacto económico en nuestro sistema sanitario.

Aunque la mayoría de los ECA muestran análisis por subgrupos sobre la eficacia de los tratamientos, interpretaciones incorrectas del análisis por subgrupos pueden afectar negativamente en la seguridad de los pacientes. En este aspecto, la combinación del esquema N+I presenta una mención especial. Esta combinación de inmunoterapia tiene asociada una elevada toxicidad y, por tanto, sólo debería ser usada cuando exista evidencia suficiente sobre un beneficio claro y relevante. En este trabajo, hemos podido observar que se ha señalado a N+I como una opción terapéutica factible en el tratamiento neoadyuvante de carcinoma de células escamosas de la cavidad oral no tratado previamente (103). También se sugirió una asociación significativa del grado de respuesta patológica al tratamiento con la presencia de células CD4 en la cohorte de pacientes tratados con N+I, pero no para el brazo de nivolumab. Tras la evaluación de estas afirmaciones con nuestra herramienta de análisis sistemático de datos de subgrupos, se determinó que no se puede aceptar de forma concluyente una asociación entre la eficacia de N+I y la expresión de CD-4 en células T por las importantes limitaciones del estudio abordado (55,103,145). De una manera similar, Gettinger *et al* relacionaron un beneficio de N+I en SG para CPNM escamoso avanzado pre-tratado, con expresión de PD-L1 <1% y una TMB ≥ 10 mt/Mb (104). Nuevamente, la aplicación de nuestra herramienta sobre análisis por subgrupos recomendó que no se debería aceptar de forma concluyente este beneficio de N+I en este contexto clínico (55,104,146). El ensayo clínico obtuvo un resultado global negativo y reclutó un insuficiente número de pacientes con estos niveles de biomarcadores.

Estos dos ejemplos demostraron que la interpretación inadecuada de análisis por subgrupos puede llevar a considerar combinaciones de fármacos con una toxicidad importante –como N+I– en contextos clínicos en los que el beneficio es, cuanto menos, dudoso. Estas situaciones no son anecdóticas, ya que el régimen N+I ya se había visto anteriormente involucrado en afirmaciones controvertidas. El ensayo CHECKMATE 067 representó gráficas del análisis por subgrupos teniendo en cuenta la presencia de mutaciones BRAF en melanoma avanzado no tratado previamente (43). Sin embargo, el

aparente beneficio en SG de la combinación de N+I frente a nivolumab para pacientes con mutaciones BRAF no fue justificado por una $p(i)$ significativa (44). En este caso, la ausencia de heterogeneidad para SG entre los subgrupos de pacientes con y sin mutaciones BRAF pudo incluso comprobarse visualmente mediante un considerable solapamiento de los intervalos de confianza. Por tanto, sólo se debió aceptar el resultado global del ECA –sin diferencia estadísticamente significativa de la combinación de inmunoterapia frente a nivolumab– y no valorar los resultados de los subgrupos. Por otra parte, no se encontraron hipótesis razonables y sólidas que respaldaran un mayor beneficio del regimen de N+I en pacientes con mutaciones BRAF. El mecanismo de acción de estos agentes inmunoterápicos no tiene relación con las mutaciones BRAF. Esto es indicativo de que las diferencias entre los subgrupos que se observaron en las gráficas podrían ser un “artefacto” por el azar o desbalanceamiento de factores. Por tanto, si se hubieran considerado válidas estas aparentes diferencias numéricas, se hubieran tratado a los pacientes con mutaciones BRAF con una combinación de fármacos de escasa tolerancia sin un beneficio real. La **Tabla 21** muestra un resumen de los datos por subgrupos según mutaciones BRAF del ensayo CHECKMATE 067.

Teniendo todo esto en cuenta, podemos afirmar que aquellas terapias con perfil negativo de seguridad sin un beneficio significativamente demostrado en una determinada subpoblación deben ser utilizadas con precaución. El uso de terapias mal toleradas podría conducir a aumentar el número de visitas hospitalarias por efectos adversos, empeorar la calidad de vida de los pacientes e incrementar los costes indirectos asociados (147). Esto haría que el balance beneficio-riesgo de los tratamientos usados fuera negativo y los pacientes se verían perjudicados. Por tanto, ECA confirmatorios de subgrupos concretos se deben desarrollar para resolver las dudas sobre el beneficio de tratamientos con considerables efectos adversos.

Tabla 21. Datos de análisis por subgrupos del estudio CHECKMATE 067.

Estudio	Población incluida en análisis por subgrupos	Grupo intervención	Grupo comparador	Variable	Subgrupo de población	Nº de pacientes	Riesgo relativo, hazard ratio (95% IC)	Interacción estadística entre subgrupos [p(i)]	Interpretación incorrecta	Interpretación adecuada	Relevancia del error de interpretación
Wolchok <i>et al</i> (ensayo CHECKMATE 067) (43)	630	Nivolumab 1 mg/kg más ipilimumab 3 mg/kg cada 3 semanas durante 2 ciclos, seguido de nivolumab 3 mg/kg cada 2 semanas	Nivolumab 3 mg/kg cada 2 semanas	Supervivencia global	Con mutaciones BRAF	201	0,70 (0,46-1,05)	0,33	Beneficio numéricamente superior en supervivencia global y supervivencia libre de progresión para la terapia en combinación frente a nivolumab en monoterapia	No se ha observado heterogeneidad entre subgrupos en supervivencia global. Se ha observado heterogeneidad baja entre subgrupos en supervivencia libre de progresión y compatible con azar o desbalanceamiento de factores	Selección de tratamiento más tóxico (nivolumab más ipilimumab) sin beneficios en supervivencia global si se consideran diferencias no reales entre subgrupos
					Sin mutaciones BRAF	429	0,89 (0,69-1,15)				
				Supervivencia libre de progresión	Con mutaciones BRAF	201	0,60 (0,43-0,86)	0,06			
					Sin mutaciones BRAF	429	0,89 (0,70-1,13)				

P(i) fue estimado usando calculadoras e información de bibliografía publicada previamente (12,33,68–72).

La conveniencia de los fármacos también puede verse afectada de forma indirecta por una aplicabilidad inadecuada de resultados por subgrupos. Esto puede observarse en el ejemplo descrito anteriormente sobre el uso de N+I para melanoma avanzado con mutaciones BRAF no tratado previamente. Si aceptáramos las aparentes diferencias gráficas –sin significación estadística– “favorables” al régimen N+I frente a nivolumab (43), aumentaría el tiempo de estancia de los pacientes en hospital de día oncológico. El tiempo necesario para la administración de N+I es mayor al tiempo requerido por nivolumab en monoterapia (148). Esto podría influir no sólo en el paciente individual, sino en el colectivo de las visitas a hospital de día. Un incremento del tiempo de administración de los tratamientos favorecería la ocupación de los sillones de administración durante más tiempo. Estas plazas de administración ambulatorias son un recurso limitado, y podría ocurrir un colapso de hospital de día. Por tanto, una mala interpretación del análisis por subgrupos podría contribuir a un incremento de la presión asistencial por el retraso de administración de tratamientos.

El desarrollo de nuevas moléculas de alto impacto fármaco-económico se ha visto incrementado considerablemente en los últimos años. De hecho, la OMS ha alertado del excesivo gasto asociado a nuevos tratamientos onco-hematológicos recientemente (26). La caducidad de las patentes es una estrategia que permite minimizar costes y favorecer la competencia de precios entre fármacos. Así, se generan oportunidades de adquisición de medicamentos de forma eficiente. Sin embargo, la existencia de interpretaciones inadecuadas de análisis por subgrupos puede dificultar o impedir la adquisición de fármacos a menor coste.

Como ya hemos comentado anteriormente, docetaxel es un anticanceroso usado en combinación con ADT en mHSPC (64). Basándose en resultados de subgrupos, algunos autores sugirieron que el esquema docetaxel más ADT aporta beneficio en SG sólo a los pacientes con elevado volumen tumoral (65). Esta aplicación del análisis por subgrupos no valoró ciertas características basales de la población –como la proporción de pacientes con terapia previa radical–, las cuales podrían influir en los resultados de los subgrupos (64,77). Además, el ECA más importante respalda el uso de docetaxel en pacientes diagnosticados de mHSPC, independientemente de la carga de la enfermedad

(66). El análisis por subgrupos presenta como limitación un error alfa elevado. La aplicación de un análisis por subgrupos requiere que existan datos consistentes entre estudios similares (55). Docetaxel ha demostrado un aumento de SG considerable en ECA sobre mHSPC (66). Por tanto, descartar a docetaxel como alternativa terapéutica para los pacientes con baja carga tumoral podría perjudicarles desde el punto de vista clínico. Además, considerar a un fármaco sin patente en vigor como alternativa terapéutica podría generar competencia de precios con otros fármacos en esta patología, como abiraterona o apalutamida (80,149). Teniendo en cuenta el impacto económico de las terapias para mHSPC, promover esta competencia de precios es esencial para reducir los costes y favorecer la sostenibilidad de los sistemas sanitarios. Los datos y representaciones de subgrupos según carga tumoral en los estudios descritos de mHSPC pueden consultarse en la **Figura 12** y la **Tabla 22**.

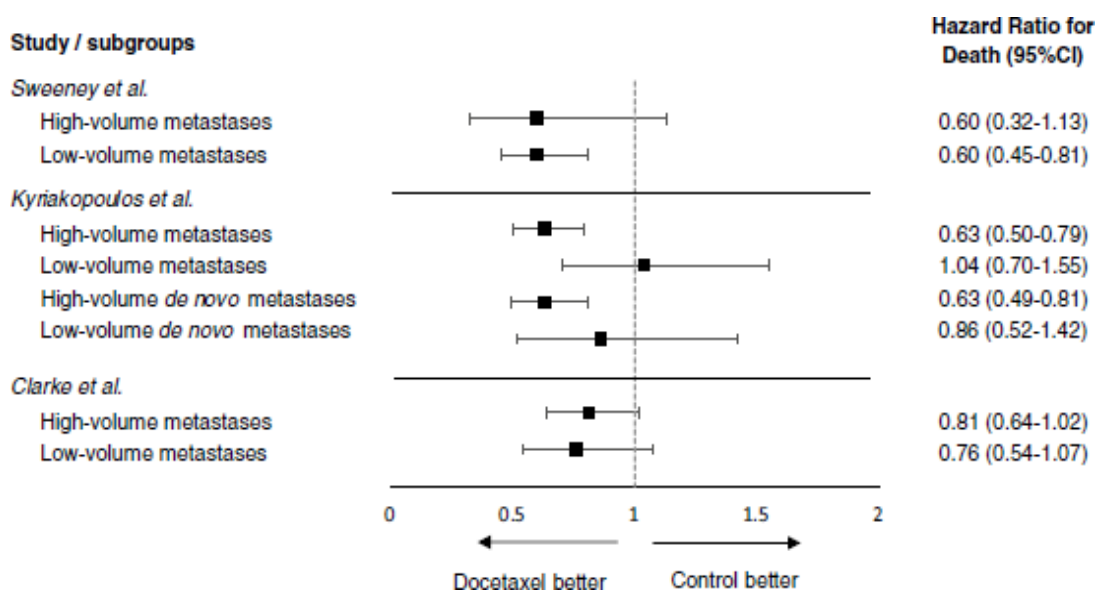


Figura 12. Representación de *forest plot* de análisis por subgrupos según el volumen de enfermedad metastásica de ECA sobre mHSPC. Tomada de Gil-Sierra *et al* (2022) (150).

Tabla 22. Datos de subgrupos según volumen de enfermedad metastásica sobre el uso de docetaxel en mHSPC.

Estudio	Población incluida en análisis por subgrupos	Intervención	Comparador	Variable	Subgrupo	Número de pacientes	Riesgo relativo, hazard ratio (95% IC)	Interacción estadística entre subgrupos [p(i)]	Interpretación incorrecta	Interpretación adecuada	Relevancia del error de interpretación
Sweeney <i>et al</i> (ensayo CHAARTED) (64)	790	Terapia de privación androgénica + docetaxel 75 mg/m ² cada 3 semanas, durante 6 ciclos	Terapia de privación androgénica	Supervivencia global	Elevado volumen metastásico	513	0,60 (0,32-1,13)	1†	Beneficio clínico superior en supervivencia global para pacientes con elevada carga tumoral tratados con terapia de privación androgénica más docetaxel según gráficas y valor de medianas	No se ha observado heterogeneidad entre subgrupos en supervivencia global.	Se descarta una alternativa terapéutica (docetaxel) con importante beneficio clínico para el subgrupo de pacientes con bajo volumen tumoral. Disminución de competencia de precios y eficiencia
					Bajo volumen metastásico	277	0,60 (0,45-0,81)				
Kyriakopoulos <i>et al</i> (actualización del ensayo CHAARTED) (65)	790	Terapia de privación androgénica + docetaxel 75 mg/m ² cada 3 semanas, durante 6 ciclos	Terapia de privación androgénica	Supervivencia global	Elevado volumen metastásico	513	0,63 (0,5-0,79)	0,03‡	El beneficio clínico de aumento en supervivencia global sólo fue demostrado en el subgrupo de pacientes con alto volumen metastasico	La proporción de pacientes con metástasis <i>de novo</i> es muy inferior en el ensayo CHAARTED en comparación con Clarke <i>et al</i> . No se encontró heterogeneidad entre subgrupos según carga tumoral en pacientes con metástasis <i>de novo</i> en estudio CHAARTED.	Se descarta una alternativa terapéutica (docetaxel) con importante beneficio clínico para el subgrupo de pacientes con bajo volumen tumoral. Disminución de competencia de precios y eficiencia
					Bajo volumen metastásico	277	1,04 (0,70-1,55)				
					Elevado volumen metastásico (metastasis <i>de novo</i>)	214	0,63 (0,49-0,81)	0,28†			
					Bajo volumen metastásico (metastasis <i>de novo</i>)	75	0,86 (0,52-1,42)				
Clarke <i>et al</i> (actualización del ensayo STAMPEDE) (66)	1086	Terapia de privación androgénica + docetaxel 75 mg/m ² cada 3 semanas, durante 6 ciclos	Terapia de privación androgénica	Supervivencia global	Elevado volumen metastásico	468	0,81 (0,64-1,02)	0,83‡	-	No se observó heterogeneidad entre los subgrupos según volumen tumoral. Docetaxel aporta beneficio en supervivencia global a los pacientes independientemente del volumen metastasico.	-
					Bajo volumen metastásico	362	0,76 (0,54-1,07)				

† P(i) fue estimado usando calculadoras e información de bibliografía publicada previamente (12,33,68-72). ‡ Datos obtenidos de los estudios.

Una mayor eficacia en un subgrupo de pacientes dentro de una patología permite a un fármaco desmarcarse de sus competidores. Esto proporciona un valor añadido a una molécula. Las compañías farmacéuticas usan frecuentemente estos datos para negociar precios más elevados de sus medicamentos con las agencias reguladoras. Algunas veces, se sugiere un mayor beneficio para un subgrupo de pacientes sin un valor de $p(i)$ estadísticamente significativo. Esto suele influir en los costes de una terapia. Un ejemplo podría ser el caso de abemaciclib para cáncer de mama avanzado con receptor hormonal positivo y ERBB2 negativo tras progresión a terapia endocrina (estudio MONARCH 2) (102). Si se tuvieran en cuenta las diferencias aparentes en SG para las subpoblaciones con resistencia primaria a terapia endocrina y/o metástasis visceral, estos datos podrían justificar (erróneamente) un mayor coste de abemaciclib. No obstante, la adecuada interpretación del análisis por subgrupos con nuestra herramienta permitió desechar ese hipotético beneficio de abemaciclib, promoviendo la competitividad de precios frente a otros inhibidores de quinasas dependientes de ciclinas (144,151,152).

El estudio fármaco-económico que hemos desarrollado sobre el uso de mepolizumab en asma eosinofílica refractaria grave mediada y no mediada por niveles elevados de IgE constituye otro ejemplo de un manejo correcto de resultados por subgrupos. En este caso, la priorización de mepolizumab en pacientes con asma eosinofílica refractaria grave no mediada por IgE y niveles plasmáticos ≥ 500 eosinófilos/ μL favorecería la sostenibilidad del sistema sanitario, pues se trata del subgrupo en el que se observó una mayor eficiencia (153). De hecho, este criterio fue usado en el IPT de mepolizumab de la Agencia Española del Medicamento (125). De esta manera, se puede comprobar que una mejor eficiencia del uso de los medicamentos puede alcanzarse mediante la evaluación del análisis por subgrupos en cualquier escenario clínico, y no sólo en patologías onco-hematológicas.

La aplicación adecuada del análisis por subgrupos también puede generar el posicionamiento racional de las alternativas terapéuticas. De hecho, el correcto uso de los datos de subgrupos y de comparaciones indirectas ajustadas permite desarrollar

estrategias de optimización. Una de ellas consiste en reservar algunos componentes de esquemas de varios fármacos con un beneficio dudoso en pacientes no tratados para líneas de tratamiento posteriores en un contexto clínico. De esta manera, los pacientes pueden disponer de más opciones terapéuticas frente a la progresión de la enfermedad.

El estudio KEYNOTE-189 analizó la combinación de pembrolizumab más quimioterapia en CPNM avanzado no tratado previamente (101). Los autores concluyeron que los pacientes con una expresión de PD-L1 mayor o igual del 50% obtuvieron un mayor beneficio de pembrolizumab asociado a quimioterapia. Sin embargo, no se demostró heterogeneidad para SG ni SLP [$p(i) > 0.1$] entre los subgrupos de pacientes con expresión de PD-L1 mayor o igual del 50% y menor del 50% (110). Asimismo, una comparación indirecta ajustada demostró ausencia de diferencia estadísticamente significativa para SLP y SG entre el esquema de pembrolizumab con quimioterapia y pembrolizumab en monoterapia en CPNM metastásico no tratado previamente y expresión de PDL-1 del 50% o superior (17,101,109). La población del ensayo KEYNOTE-189 habían sido tratados con terapia basada en platino con pemetrexed. En el estudio KEYNOTE-024 se evaluó la eficacia de pembrolizumab frente a quimioterapia (incluyendo pemetrexed y otros fármacos) en CPNM metastásico sin tratamiento previo y expresión de PD-L1 mayor o igual a 50% (17). En este ensayo, se encontró heterogeneidad para SLP entre la subpoblación que recibió quimioterapia basada en platino con pemetrexed y la subpoblación tratada con esquemas sin pemetrexed. Por ello, fue necesario desarrollar una nueva comparación indirecta ajustada para SG y SLP como análisis de sensibilidad. Se usaron los resultados de eficacia para SG y SLP de los subgrupos del estudio KEYNOTE-189 con una expresión de PD-L1 igual al 50% o mayor. Del estudio KEYNOTE-024, se seleccionaron los datos de SLP para el subgrupo de pacientes tratado con quimioterapia basada en platino más pemetrexed, y los resultados globales de SG. En esta nueva comparación indirecta ajustada, se observaron diferencias estadísticamente significativas para SLP a favor de pembrolizumab con quimioterapia y pemetrexed, pero sin beneficio para SG.

El posicionamiento terapéutico de los diferentes esquemas en un contexto clínico suele ser complejo. En este caso de pembrolizumab más quimioterapia debería tenerse en

cuenta el dudoso beneficio en SG como primera línea de tratamiento, las desventajas de seguridad de la combinación de fármacos, las limitaciones del análisis por subgrupos y sus comparaciones indirectas, y el importante beneficio de quimioterapia basada en platino con pemetrexed en CPNM metastásico tratado previamente (101). Por ello, un correcto manejo del análisis por subgrupos permitió recomendar objetivamente el uso de pembrolizumab en monoterapia para CPNM metastásico no tratado previamente y con expresión de PD-L1 igual o superior al 50%. De esta manera, se reservaría otro buen esquema –quimioterapia basada en platino con pemetrexed– frente a la progresión tumoral en estos pacientes.

Como ya hemos comprobado, los resultados de subgrupos pueden ser empleados para llevar a cabo comparaciones entre fármacos o posicionar las diferentes alternativas terapéuticas en una enfermedad. El uso de un subgrupo de pacientes como parte del conjunto de datos de una comparación indirecta o NMA puede ser considerada como una limitación metodológica que reduce la credibilidad de los resultados obtenidos. Las comparaciones indirectas y NMAs deben desarrollarse evitando mezclar poblaciones con características diferentes. Estos factores podrían tener un impacto en el pronóstico de los pacientes y evolución de la enfermedad (154). Antes de comparar los datos, se debe considerar si el resultado de una subpoblación en un ECA puede presentar peor o mejor pronóstico que el total de la población incluida en otro ECA. Por ello, la combinación de datos de poblaciones globales de ECA con datos de subgrupos podría asociarse con un mayor grado de incertidumbre en la toma de decisiones de la práctica médica.

No obstante, existen escenarios clínicos en los que algunos medicamentos sólo disponen de datos en determinados subgrupos dentro de un ECA (155). Para intentar minimizar sesgos, se puede comprobar la heterogeneidad entre subgrupos según $p(i)$ antes de seleccionar los datos para un estudio comparativo. Un NMA sobre eficacia de fármacos antiangiogénicos en primera línea de tratamiento para cáncer renal usó esta estrategia (156). Algunos de los antiangiogénicos incluidos en el estudio sólo habían sido analizados en ECA que reclutaron tanto pacientes diagnosticados de cáncer renal no tratados previamente como población pretratada. No se encontró heterogeneidad entre estas dos subpoblaciones dentro de los ECA valorados. Por tanto, se decidió usar los

datos globales de los ECA con ambos subgrupos para combinarlos en el NMA con el resto de ECA que incluyeron sólo pacientes naive. Otra estrategia hubiera sido seleccionar sólo los datos de los subgrupos no tratados, asumiendo las limitaciones de tamaños muestrales pequeños. Un análisis de sensibilidad con las dos estrategias de selección de datos citadas podría ser de utilidad en la toma de decisiones clínicas. Es importante reseñar que, si se hubiera observado heterogeneidad en los ECA que reclutaron pacientes con cáncer renal no tratado y población que había recibido terapias previas, sólo se podrían haber seleccionado los datos de la población no tratada para elaborar el NMA. Todas estas consideraciones deberían ser tenidas en cuenta antes de desarrollar comparaciones entre alternativas terapéuticas para que la credibilidad de los resultados sea mínimamente fiable.

En este trabajo se ha demostrado la importante –y creciente– influencia que el análisis por subgrupos puede tener en la toma de decisiones fármaco-económicas, clínicas, y en el desarrollo de estudios posteriores. Los resultados del análisis por subgrupos pueden tener un notable impacto en la evaluación de la eficacia, seguridad, adecuación, eficiencia y posicionamiento de las terapias. Este impacto puede verse incrementado de forma exponencial en determinadas situaciones de emergencia, como la crisis sanitaria COVID-19. Este tipo de escenarios exigen que se aceleren las etapas de desarrollo de fármacos para hacer frente a situaciones de extrema necesidad, lo que lamentablemente también puede contribuir a realizar afirmaciones prematuras sobre estas nuevas moléculas. Por ello, el desarrollo de herramientas objetivas y sistemáticas que ayuden al abordaje del análisis por subgrupos puede ser de gran ayuda para los profesionales sanitarios (4,55). La **Figura 13** muestra un algoritmo que puede facilitar la aplicación del análisis por subgrupos.

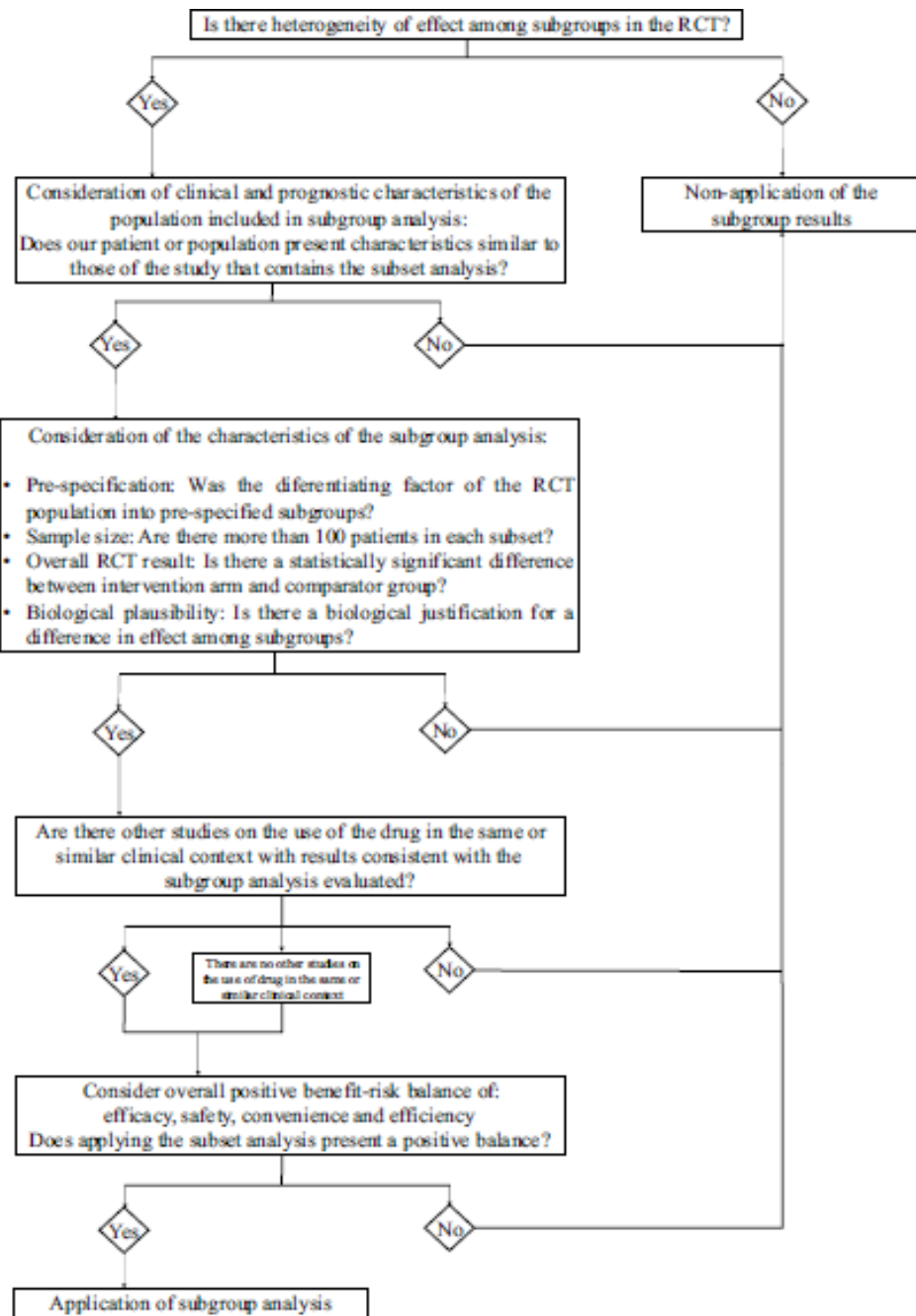


Figura 13. Algoritmo de decisión para la aplicación del análisis por subgrupos. Tomada de Gil-Sierra *et al* (2022) (150).

Finalmente, podemos afirmar que nuestro trabajo ha permitido desarrollar una metodología sistemática y ordenada sobre el análisis por subgrupos que tiene en cuenta unos criterios definidos para favorecer una toma de decisiones racional, evitar importantes repercusiones negativas en los pacientes o sistemas de salud, y la hemos aplicado a numerosos casos de elevado impacto clínico y fármaco-económico.

5. CONCLUSIONES

5. CONCLUSIONES

Las conclusiones que se pueden extraer de esta investigación son las siguientes:

1. La herramienta sistemática desarrollada para el abordaje sobre el análisis por subgrupos –con un diseño y validación que tuvieron en cuenta las limitaciones de estos resultados– obtuvo valoraciones positivas por los usuarios, mostrándose útil y fácil de manejar usar. Esta herramienta se presenta como una metodología con la que diferentes profesionales sanitarios pueden entablar discusiones objetivas en la toma de decisiones terapéuticas que involucran a análisis por subgrupos.

2. Se llevó a cabo la aplicación de la herramienta desarrollada en varios escenarios de elevado impacto clínico, fármaco-económico y social. Se realizaron estudios con revisiones sistemáticas y uso de nuestra metodología sobre análisis por subgrupos tanto de fármacos onco-hematológicos como de otros contextos clínicos (COVID-19), evitando interpretaciones inadecuadas de estos datos. La aplicación de nuestra herramienta sobre publicaciones concretas también mostró las ventajas de nuestro sistema de evaluación de resultados de subgrupos a la hora de interpretar datos. El uso de la herramienta podría evitar que afirmaciones erróneas sobre los resultados de subgrupos tengan un impacto negativo en la eficacia, seguridad, adecuación y posicionamiento terapéutico de los fármacos. Además, esta metodología podría minimizar el sesgo en la selección de datos para la realización de nuevos estudios.

3. El estudio fármaco-económico sobre mepolizumab en asma eosinofílica refractaria grave demostró que una evaluación fármaco-económica adecuada del análisis por subgrupos puede contribuir a la toma de decisiones eficientes y a la sostenibilidad del sistema sanitario, llegando incluso a considerar los subgrupos como criterio pivotal en documentos institucionales estatales.

4. El papel del farmacéutico durante el desarrollo y aplicación de la herramienta ha sido esencial. Las posibilidades de aplicación de nuestra herramienta son muy elevadas debido al carácter transversal de la misma. De esta manera, nuestra metodología puede tener un considerable impacto clínico y farmacoeconómico tanto a nivel hospitalario como de atención primaria, por lo que cualquier paciente o sistema sanitario podría verse beneficiado.

6. BIBLIOGRAFÍA

6. BIBLIOGRAFÍA

1. Wang R, Lagakos SW, Ware JH, Hunter DJ, Drazen JM. Statistics in medicine--reporting of subgroup analyses in clinical trials. *N Engl J Med.* 2007;357(21):2189-94. Doi: 10.1056/NEJMSr077003.
2. Sun X, Briel M, Walter SD, Guyatt GH. Is a subgroup effect believable? Updating criteria to evaluate the credibility of subgroup analyses. *BMJ.* 2010;340:c117. Doi: 10.1136/bmj.c117.
3. Vidic A, Chibnall JT, Goparaju N, Hauptman PJ. Subgroup analyses of randomized clinical trials in heart failure: facts and numbers. *ESC Heart Fail.* 2016;3(3):152-7. Doi: 10.1002/ehf2.12093
4. Sun X, Ioannidis JPA, Agoritsas T, Alba AC, Guyatt G. How to use a subgroup analysis: users' guide to the medical literature. *JAMA.* 2014;311(4):405-11. Doi: 10.1001/jama.2013.285063.
5. Brookes ST, Whitley E, Peters TJ, Mulheran PA, Egger M, Davey Smith G. Subgroup analyses in randomised controlled trials: quantifying the risks of false-positives and false-negatives. *Health Technol Assess.* 2001;5(33):56. Doi: 10.3310/hta5330.
6. Báez-Gutiérrez N, Rodríguez-Ramallo H, Flores-Moreno S, Abdel-Kader Martín L. Subgroup analysis in haematologic malignancies phase III clinical trials: A systematic review. *Br J Clin Pharmacol.* 2021 Jul;87(7):2635-2644. Doi: 10.1111/bcp.14689
7. Mateos MV, Dimopoulos MA, Cavo M, Suzuki K, Jakubowiak A, Knop S, et al. Daratumumab plus bortezomib, melphalan, and prednisone for untreated myeloma. *N Engl J Med.* 2018;378(6):518-28. Doi: 10.1056/NEJMoa1714678.
8. Wallach JD, Sullivan PG, Trepanowski JF, Sainani KL, Steyerberg EW, Ioannidis JPA. Evaluation of evidence of statistical support and corroboration of subgroup claims in randomized clinical trials. *JAMA Intern Med.* 2017;177(4):554-60. Doi: 10.1001/jamainternmed.2016.9125.
9. Matthews JNS, Altman DG. Statistics notes. Interaction 2: compare effect sizes not P values. *BMJ.* 1996;313(7060):808. Doi: 10.1136/bmj.313.7060.808.

10. Matthews JN, Altman DG. Interaction 3: How to examine heterogeneity. *BMJ*. 1996;313(7061):862. Doi: 10.1136/bmj.313.7061.862.
11. Barraclough H, Govindan R. Biostatistics primer: what a clinician ought to know: subgroup analyses. *J Thorac Oncol*. 2010;5(5):741-6. Doi: 10.1097/JTO.0b013e3181d9009e
12. Altman DG, Matthews JN. Statistics notes. Interaction 1: Heterogeneity of effects. *BMJ*. 1996;313(7055):486. Doi: 10.1136/bmj.313.7055.486
13. Perneger TV. What's wrong with Bonferroni adjustments. *BMJ*. 1998;316(7139):1236-8. Doi: 10.1136/bmj.316.7139.1236
14. Lagakos SW. The challenge of subgroup analyses — reporting without distorting. *N Engl J Med*. 2006;354(16):1667-9. Doi: 10.1056/NEJMp068070
15. Antonia SJ, Villegas A, Daniel D, Vicente D, Murakami S, Hui R, et al. Overall survival with durvalumab after chemoradiotherapy in stage III NSCLC. *N Engl J Med*. 2018;379(24):2342-50. Doi: 10.1056/NEJMoa1809697
16. Herbst RS, Baas P, Kim DW, Felip E, Pérez-Gracia JL, Han JY, et al. Pembrolizumab versus docetaxel for previously treated, PD-L1-positive, advanced non-small-cell lung cancer (KEYNOTE-010): a randomised controlled trial. *Lancet*. 2016;387(10027):1540-50. Doi: 10.1016/S0140-6736(15)01281-7
17. Reck M, Rodríguez-Abreu D, Robinson AG, Hui R, Csőszi T, Fülöp A, et al. Pembrolizumab versus chemotherapy for PD-L1-positive non-small-cell lung cancer. *N Engl J Med*. 10 de noviembre de 2016;375(19):1823-33. Doi: 10.1056/NEJMoa1606774
18. Song F, Bachmann MO. Cumulative subgroup analysis to reduce waste in clinical research for individualised medicine. *BMC Med*. 2016;14(1):197. Doi:10.1186/s12916-016-0744-x
19. Kaiser LD. Stratification of randomization is not required for a pre-specified subgroup analysis. *Pharm Stat*. 2013;12(1):43-7. Doi: 10.1002/pst.1550
20. Sun X, Briel M, Busse JW, You JJ, Akl EA, Mejza F, et al. Credibility of claims of subgroup effects in randomised controlled trials: systematic review. *BMJ*. 2012;344:e1553. doi: 10.1136/bmj.e1553

21. Gabler NB, Duan N, Ranases E, Suttner L, Ciarametaro M, Cooney E, et al. No improvement in the reporting of clinical trial subgroup effects in high-impact general medical journals. *Trials*. 2016;17(1):320. Doi: 10.1186/s13063-016-1447-5
22. Anonymous. ICH E3 structure and content of clinical study reports - Scientific guideline [Internet]. European Medicines Agency. 2018 [citado 12 de diciembre de 2022]. Disponible en: <https://www.ema.europa.eu/en/ich-e3-structure-content-clinical-study-reports>
23. Petticrew M, Tugwell P, Kristjansson E, Oliver S, Ueffing E, Welch V. Damned if you do, damned if you don't: subgroup analysis and equity. *J Epidemiol Community Health*. 2012;66(1):95-8. Doi: 10.1136/jech.2010.121095
24. Rothwell PM. Treating individuals 2. Subgroup analysis in randomised controlled trials: importance, indications, and interpretation. *Lancet*. 2005;365(9454):176-86. Doi: 10.1016/S0140-6736(05)17709-5
25. Zhang S, Liang F, Li W, Hu X. Subgroup Analyses in Reporting of Phase III Clinical Trials in Solid Tumors. *J Clin Oncol*. 2015;33(15):1697-702. Doi: 10.1200/JCO.2014.59.8862
26. World Health Organization. Technical report: pricing of cancer medicines and its impacts: a comprehensive technical report for the World Health Assembly Resolution 70.12: operative paragraph 2.9 on pricing approaches and their impacts on availability and affordability of medicines for the prevention and treatment of cancer [Internet]. World Health Organization. 2018 [citado 12 de diciembre de 2022]. Disponible en: <https://apps.who.int/iris/handle/10665/277190>
27. Hussain M, Fizazi K, Saad F, Rathenborg P, Shore N, Ferreira U, et al. Enzalutamide in men with nonmetastatic, castration-resistant prostate cancer. *N Engl J Med*. 2018;378(26):2465-74. Doi: 10.1056/NEJMoa1800536
28. Tombal B, Saad F, Penson D, Hussain M, Sternberg CN, Morlock R, et al. Patient-reported outcomes following enzalutamide or placebo in men with non-metastatic, castration-resistant prostate cancer (PROSPER): a multicentre, randomised, double-blind, phase 3 trial. *Lancet Oncol*. 2019;20(4):556-69. Doi: 10.1016/S1470-2045(18)30898-2
29. Smith MR, Saad F, Chowdhury S, Oudard S, Hadaschik BA, Graff JN, et al. Apalutamide treatment and metastasis-free survival in prostate cancer. *N Engl J Med*. 2018;378(15):1408-18. Doi: 10.1056/NEJMoa1715546

30. Pavord ID, Korn S, Howarth P, Bleecker ER, Buhl R, Keene ON, et al. Mepolizumab for severe eosinophilic asthma (DREAM): a multicentre, double-blind, placebo-controlled trial. *Lancet*. 2012;380(9842):651-9. Doi: 10.1016/S0140-6736(12)60988-X
31. Wang Y, Zhang D, Du G, Du R, Zhao J, Jin Y, et al. Remdesivir in adults with severe COVID-19: a randomised, double-blind, placebo-controlled, multicentre trial. *Lancet*. 2020;395(10236):1569-78. Doi: 10.1016/S0140-6736(20)31022-9
32. Hirji KF, Fagerland MW. Outcome based subgroup analysis: a neglected concern. *Trials* 2009;10:33. Doi: 10.1186/1745-6215-10-33.
33. Altman DG, Bland JM. Interaction revisited: the difference between two estimates. *BMJ*. 2003;326(7382):219. Doi: 10.1136/bmj.326.7382.219.
34. Fuchs CS, Tomasek J, Yong CJ, Dumitru F, Passalacqua R, Goswami C, et al. Ramucirumab monotherapy for previously treated advanced gastric or gastro-oesophageal junction adenocarcinoma (REGARD): an international, randomised, multicentre, placebo-controlled, phase 3 trial. *Lancet*. 2014;383(9911):31-9. Doi: 10.1016/S0140-6736(13)61719-5
35. Wilke H, Muro K, Van Cutsem E, Oh SC, Bodoky G, Shimada Y, et al. Ramucirumab plus paclitaxel versus placebo plus paclitaxel in patients with previously treated advanced gastric or gastro-oesophageal junction adenocarcinoma (RAINBOW): a double-blind, randomised phase 3 trial. *Lancet Oncol*. 2014;15(11):1224-35. Doi: 10.1016/S1470-2045(14)70420-6
36. Motzer RJ, Escudier B, McDermott DF, George S, Hammers HJ, Srinivas S, et al. Nivolumab versus everolimus in advanced renal-cell carcinoma. *N Engl J Med*. 2015;373(19):1803-13. Doi: 10.1056/NEJMoa1510665.
37. Ortega HG, Liu MC, Pavord ID, Brusselle GG, FitzGerald JM, Chetta A, et al. Mepolizumab treatment in patients with severe eosinophilic asthma. *N Engl J Med*. 2014;371(13):1198-207. Doi: 10.1056/NEJMoa1403290.
38. Brouwers MC, Kho ME, Browman GP, Burgers JS, Cluzeau F, Feder G, et al. AGREE II: advancing guideline development, reporting, and evaluation in health care. *Prev Med*. 2010;51(5):421-4. Doi: 10.1016/j.ypmed.2010.08.005
39. Altman DG. *Practical Statistics for Medical Research*. CRC Press; 1990.

40. Kent DM, Nelson J, Dahabreh IJ, Rothwell PM, Altman DG, Hayward RA. Risk and treatment effect heterogeneity: re-analysis of individual participant data from 32 large clinical trials. *Int J Epidemiol.* 2016;45(6):2075-88. Doi: 10.1093/ije/dyw118.
41. Grupo GÉNESIS de Sociedad Española de Farmacia Hospitalaria. Programa Madre [Internet]. Sociedad Española de Farmacia Hospitalaria. 2005 [citado 12 de diciembre de 2022]. Disponible en: <https://gruposdetrabajo.sefh.es/genesis/index.php/11-programa-madre>
42. Grupo GHEMA de Sociedad Andaluza de Farmacéuticos Hospitalarios [Internet]. 2022 [citado 12 de diciembre de 2022]. Disponible en: <https://safh.org/ghema/>
43. Wolchok JD, Chiarion-Sileni V, Gonzalez R, Rutkowski P, Grob JJ, Cowey CL, et al. Overall survival with combined nivolumab and ipilimumab in advanced melanoma. *N Engl J Med.* 2017;377(14):1345-56. Doi: 10.1056/NEJMoa1709684
44. Alegre-del-Rey EJ, de la Nogal Fernández B, Briceño-Casado P. Nivolumab and ipilimumab in advanced melanoma. *N Engl J Med.* 2017;377(25):2503. Doi:10.1056/NEJMc1714339
45. Facon T, Kumar S, Plesner T, Orlowski RZ, Moreau P, Bahlis N, et al. Daratumumab plus lenalidomide and dexamethasone for untreated myeloma. *N Engl J Med.* 2019;380(22):2104-15. Doi: 10.1056/NEJMoa1817249
46. Agencia Española de Medicamentos y Productos Sanitarios. Ficha técnica Darzalex 20 mg/ml concentrado para solución para perfusión [Internet]. Agencia Española de Medicamentos y Productos Sanitarios. 2016 [citado 12 de diciembre de 2022]. Disponible en: https://cima.aemps.es/cima/dochtml/ft/1161101001/FT_1161101001.html
47. Agencia Española de Medicamentos y Productos Sanitarios. Ficha técnica Docetaxel Accord 20 mg/ml concentrado para solución para perfusión [Internet]. Agencia Española de Medicamentos y Productos Sanitarios. 2012 [citado 12 de diciembre de 2022]. Disponible en: https://cima.aemps.es/cima/dochtml/ft/12769001/FT_12769001.html
48. Agencia Española de Medicamentos y Productos Sanitarios. Ficha técnica Veklury 100 mg polvo para concentrado para solución para perfusión [Internet]. Agencia Española de Medicamentos y Productos Sanitarios. 2020 [citado 12 de diciembre de 2022]. Disponible en: https://cima.aemps.es/cima/pdfs/ft/1201459002/FT_1201459002.pdf

49. Palumbo A, Anderson K. Multiple myeloma. *N Engl J Med.* 2011;364(11):1046-60. Doi: 10.1056/NEJMra1011442
50. San Miguel JF, Schlag R, Khuageva NK, Dimopoulos MA, Shpilberg O, Kropff M, et al. Bortezomib plus melphalan and prednisone for initial treatment of multiple myeloma. *N Engl J Med.* 2008;359(9):906-17. Doi: 10.1056/NEJMoa0801479
51. Benboubker L, Dimopoulos MA, Dispenzieri A, Catalano J, Belch AR, Cavo M, et al. Lenalidomide and dexamethasone in transplant-ineligible patients with myeloma. *N Engl J Med.* 2014;371(10):906-17. Doi: 10.1056/NEJMoa1402551
52. Mateos MV, Cavo M, Blade J, Dimopoulos MA, Suzuki K, Jakubowiak A, et al. Overall survival with daratumumab, bortezomib, melphalan, and prednisone in newly diagnosed multiple myeloma (ALCYONE): a randomised, open-label, phase 3 trial. *Lancet.* 2020;395(10218):132-41. Doi: 10.1016/S0140-6736(19)32956-3
53. Methley AM, Campbell S, Chew-Graham C, McNally R, Cheraghi-Sohi S. PICO, PICOS and SPIDER: a comparison study of specificity and sensitivity in three search tools for qualitative systematic reviews. *BMC Health Serv Res.* 2014;14:579. Doi: 10.1186/s12913-014-0579-0
54. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med.* 2009;6(7):e1000097. Doi: 10.1371/journal.pmed.1000097
55. Gil-Sierra MD, Fénix-Caballero S, Abdel Kader-Martin L, Fraga-Fuentes MD, Sánchez-Hidalgo M, Alarcón de la Lastra-Romero C, et al. Checklist for clinical applicability of subgroup analysis. *J Clin Pharm Ther.* 2020;45(3):530-8. Doi: 10.1111/jcpt.13102
56. International Myeloma Working Group. IMWG [Internet]. IMWG. 1990 [citado 12 de diciembre de 2019]. Disponible en: <http://imwg.myeloma.org/>
57. Yan X, Clemens PL, Puchalski T, Lonial S, Lokhorst H, Voorhees PM, et al. Influence of disease and patient characteristics on daratumumab exposure and clinical outcomes in relapsed or refractory multiple myeloma. *Clin Pharmacokinet.* 2018;57(4):529-38. Doi: 10.1007/s40262-017-0598-1

58. Bröckelmann PJ, McMullen S, Wilson JB, Mueller K, Goring S, Stamatoullas A, et al. Patient and physician preferences for first-line treatment of classical Hodgkin lymphoma in Germany, France and the United Kingdom. *Br J Haematol.* 2019;184(2):202-14. Doi: 10.1111/bjh.15566.
59. Laryionava K, Winkler EC. Patients' preferences in non-curable cancer disease. *Oncol Res Treat.* 2019;42(1-2):31-4. Doi: 10.1159/000496120
60. Kastbom L, Milberg A, Karlsson M. A good death from the perspective of palliative cancer patients. *Support Care Cancer.* 2017;25(3):933-9. Doi: 10.1007/s00520-016-3483-9
61. Li T, Pei X, Chen X, Zhang S. Identifying end-of-life preferences among Chinese patients with cancer using the heart to heart card game. *Am J Hosp Palliat Care.* 2021;38(1):62-7. Doi: 10.1177/1049909120917361
62. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. *CA Cancer J Clin.* 2020;70(1):7-30. Doi: 10.3322/caac.21590
63. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2018;68(6):394-424. Doi: 10.3322/caac.21492.
64. Sweeney CJ, Chen YH, Carducci M, Liu G, Jarrard DF, Eisenberger M, et al. Chemohormonal therapy in metastatic hormone-sensitive prostate cancer. *N Engl J Med.* 2015;373(8):737-46. Doi: 10.1056/NEJMoa1503747
65. Kyriakopoulos CE, Chen YH, Carducci MA, Liu G, Jarrard DF, Hahn NM, et al. Chemohormonal therapy in metastatic hormone-sensitive prostate cancer: long-term survival analysis of the randomized phase III E3805 CHARTED trial. *J Clin Oncol.* 2018;36(11):1080-7. Doi: 10.1200/JCO.2017.75.3657
66. Clarke NW, Ali A, Ingleby FC, Hoyle A, Amos CL, Attard G, et al. Addition of docetaxel to hormonal therapy in low- and high-burden metastatic hormone sensitive prostate cancer: long-term survival results from the STAMPEDE trial. *Ann Oncol.* 2019;30(12):1992-2003. Doi: 10.1093/annonc/mdz396

67. Fizazi K, Tran N, Fein L, Matsubara N, Rodriguez-Antolin A, Alekseev BY, et al. Abiraterone plus prednisone in metastatic castration-sensitive prostate cancer. *N Engl J Med.* 2017;377(4):352-60. Doi: 10.1056/NEJMoa1704174
68. Hancock MJ, Kjaer P, Korsholm L, Kent P. Interpretation of subgroup effects in published trials. *Phys Ther.* 2013;93(6):852-9. Doi: 10.2522/ptj.20120296
69. Parmar MK, Torri V, Stewart L. Extracting summary statistics to perform meta-analyses of the published literature for survival endpoints. *Stat Med.* 1998;17(24):2815-34. Doi: 10.1002/(sici)1097-0258(19981230)17:24<2815::aid-sim110>3.0.co;2-8.
70. Tierney JF, Stewart LA, Ghersi D, Burdett S, Sydes MR. Practical methods for incorporating summary time-to-event data into meta-analysis. *Trials.* 2007;8:16. Doi: 10.1186/1745-6215-8-16
71. Vale CL, Tierney JF, Stewart LA. Effects of adjusting for censoring on meta-analyses of time-to-event outcomes. *Int J Epidemiol.* 2002;31(1):107-11. Doi: 10.1093/ije/31.1.107.
72. Primo J, Escrig J. Utilidades: calculadoras y programas [Internet]. Grupo GENESIS de la Sociedad Española de Farmacia Hospitalaria. 2016 [citado 12 de diciembre de 2019]. Disponible en: https://gruposdetrabajo.sefh.es/genesis/index.php?option=com_content&view=article&id=46&Itemid=42
73. James ND, Sydes MR, Clarke NW, Mason MD, Dearnaley DP, Spears MR, et al. Addition of docetaxel, zoledronic acid, or both to first-line long-term hormone therapy in prostate cancer (STAMPEDE): survival results from an adaptive, multiarm, multistage, platform randomised controlled trial. *Lancet.* 2016;387(10024):1163-77. Doi: 10.1016/S0140-6736(15)01037-5
74. Eisenberger MA, Blumenstein BA, Crawford ED, Miller G, McLeod DG, Loehrer PJ, et al. Bilateral orchiectomy with or without flutamide for metastatic prostate cancer. *N Engl J Med.* 1998;339(15):1036-42. Doi: 10.1056/NEJM199810083391504
75. Crawford ED, Eisenberger MA, McLeod DG, Spaulding JT, Benson R, Dorr FA, et al. A controlled trial of leuprolide with and without flutamide in prostatic carcinoma. *N Engl J Med.* 1989;321(7):419-24. Doi: 10.1056/NEJM198908173210702

76. Glass TR, Tangen CM, Crawford ED, Thompson I. Metastatic carcinoma of the prostate: identifying prognostic groups using recursive partitioning. *J Urol*. 2003;169(1):164-9. Doi: 10.1097/01.ju.0000042482.18153.30
77. Gravis G, Fizazi K, Joly F, Oudard S, Priou F, Esterni B, et al. Androgen-deprivation therapy alone or with docetaxel in non-castrate metastatic prostate cancer (GETUG-AFU 15): a randomised, open-label, phase 3 trial. *Lancet Oncol*. 2013;14(2):149-58. Doi: 10.1016/S1470-2045(12)70560-0
78. Gravis G, Boher JM, Joly F, Soulié M, Albiges L, Priou F, et al. Androgen deprivation therapy (ADT) plus docetaxel versus ADT alone in metastatic non castrate prostate cancer: impact of metastatic burden and long-term survival analysis of the randomized phase 3 GETUG-AFU15 trial. *Eur Urol*. 2016;70(2):256-62. Doi: 10.1016/j.eururo.2015.11.005
79. Kim N Chi, Neeraj Agarwal, Anders Bjartell, Byung Ha Chung, Andrea J Pereira de Santana Gomes, Robert Given. Apalutamide for metastatic, castration-sensitive prostate cancer. *N Engl J Med*. 2019;381(1):13-24. Doi: 10.1056/NEJMoa1903307.
80. Fizazi K, Tran N, Fein L, Matsubara N, Rodriguez-Antolin A, Alekseev BY, et al. Abiraterone acetate plus prednisone in patients with newly diagnosed high-risk metastatic castration-sensitive prostate cancer (LATITUDE): final overall survival analysis of a randomised, double-blind, phase 3 trial. *Lancet Oncol*. 2019;20(5):686-700. Doi: 10.1016/S1470-2045(19)30082-8.
81. Finianos A, Gupta K, Clark B, Simmens SJ, Aragon-Ching JB. Characterization of differences between prostate cancer patients presenting with de novo versus primary progressive metastatic disease. *Clin Genitourin Cancer*. 2017; S1558-7673(17)30247-1. Doi: 10.1016/j.clgc.2017.08.006
82. Sydes MR, Spears MR, Mason MD, Clarke NW, Dearnaley DP, de Bono JS, et al. Adding abiraterone or docetaxel to long-term hormone therapy for prostate cancer: directly randomised data from the STAMPEDE multi-arm, multi-stage platform protocol. *Ann Oncol*. 2018;29(5):1235-48. Doi: 10.1093/annonc/mdy072.
83. Johns Hopkins Coronavirus Resource Center. COVID-19 Map [Internet]. 2022 [citado 12 de diciembre de 2022]. Disponible en: <https://coronavirus.jhu.edu/map.html>

84. Potere N, Di Nisio M, Rizzo G, La Vella M, Polilli E, Agostinone A, et al. Low-dose subcutaneous tocilizumab to prevent disease progression in patients with moderate COVID-19 pneumonia and hyperinflammation. *Int J Infect Dis.* 2020; 100:421-424. Doi: 10.1016/j.ijid.2020.07.078
85. Cao B, Wang Y, Wen D, Liu W, Wang J, Fan G, et al. A trial of lopinavir-ritonavir in adults hospitalized with severe Covid-19. *N Engl J Med.* 2020;382(19):1787-99. Doi: 10.1056/NEJMoa2001282.
86. Horby P, Lim WS, Emberson JR, Mafham M, Bell JL, Linsell L, et al. Dexamethasone in hospitalized patients with Covid-19. *N Engl J Med.* 2021;384(8):693-704. Doi: 10.1056/NEJMoa2021436.
87. Siemieniuk RA, Bartoszko JJ, Ge L, Zeraatkar D, Izcovich A, Pardo-Hernandez H, et al. Drug treatments for Covid-19: living systematic review and network meta-analysis. *BMJ.* 2020;370:m2980. Doi: 10.1136/bmj.m2980
88. Baden LR, El Sahly HM, Essink B, Kotloff K, Frey S, Novak R, et al. Efficacy and safety of the mRNA-1273 SARS-CoV-2 vaccine. *N Engl J Med.* 2021;384(5):403-16. Doi: 10.1056/NEJMoa2035389
89. Polack FP, Thomas SJ, Kitchin N, Absalon J, Gurtman A, Lockhart S, et al. Safety and efficacy of the BNT162b2 mRNA Covid-19 vaccine. *N Engl J Med.* 2020;383(27):2603-15. Doi: 10.1056/NEJMoa2034577
90. Gillenwater S, Rahaghi F, Hadeh A. Remdesivir for the treatment of Covid-19 - preliminary report. *N Engl J Med.* 2020; 383(10):992. Doi: 10.1056/NEJMc2022236
91. Goldman JD, Lye DCB, Hui DS, Marks KM, Bruno R, Montejano R, et al. Remdesivir for 5 or 10 days in patients with severe Covid-19. *N Engl J Med.* 2020; 383(19):1827-1837. Doi: 10.1056/NEJMoa2015301.
92. Spinner CD, Gottlieb RL, Criner GJ, Arribas López JR, Cattelan AM, Soriano Viladomiu A, et al. Effect of remdesivir vs standard care on clinical status at 11 days in patients with moderate COVID-19: a randomized clinical trial. *JAMA.* 2020;324(11):1048-57. Doi: 10.1001/jama.2020.16349.

93. Beigel JH, Tomashek KM, Dodd LE, Mehta AK, Zingman BS, Kalil AC, et al. Remdesivir for the treatment of Covid-19 - final report. *N Engl J Med.* 2020;383(19):1813-26. Doi: 10.1056/NEJMoa2007764.
94. Kalil AC, Patterson TF, Mehta AK, Tomashek KM, Wolfe CR, Ghazaryan V, et al. Baricitinib plus remdesivir for hospitalized adults with Covid-19. *N Engl J Med.* 2021;384(9):795-807. Doi: 10.1056/NEJMoa2031994.
95. Barratt-Due A, Olsen IC, Nezvalova-Henriksen K, Kåsine T, Lund-Johansen F, Hoel H, et al. Evaluation of the effects of remdesivir and hydroxychloroquine on viral clearance in Covid-19: a randomized trial. *Ann Intern Med.* 2021;174(9):1261-9. Doi: 10.7326/M21-0653
96. European Medicines Agency. Veklury [Internet]. European Medicines Agency. 2020 [citado 17 de diciembre de 2022]. Disponible en: <https://www.ema.europa.eu/en/medicines/human/EPAR/veklury>
97. Muthuri SG, Venkatesan S, Myles PR, Leonardi-Bee J, Al Khuwaitir TSA, Al Mamun A, et al. Effectiveness of neuraminidase inhibitors in reducing mortality in patients admitted to hospital with influenza A H1N1pdm09 virus infection: a meta-analysis of individual participant data. *Lancet Respir Med.* 2014;2(5):395-404. Doi: 10.1016/S2213-2600(14)70041-4.
98. Pan H, Peto R, Henao-Restrepo AM, Preziosi MP, Sathiyamoorthy V, Quarraisha Abdool K, et al. Repurposed antiviral drugs for Covid-19 - interim WHO Solidarity trial results. *N Engl J Med.* 2021;384(6):497-511. Doi: 10.1056/NEJMoa2023184
99. Alia E, Grant-Kels JM. Does hydroxychloroquine combat Covid-19? A timeline of evidence. *J Am Acad Dermatol.* 2020; 83(1):e33-e34. Doi: 10.1016/j.jaad.2020.04.031.
100. Hill A, Wang J, Levi J, Heath K, Fortunak J. Minimum costs to manufacture new treatments for Covid-19. *J Virus Erad.* 2020;6(2):61-9. Doi: 10.1016/S2055-6640(20)30018-2.
101. Gandhi L, Rodríguez-Abreu D, Gadgeel S, Esteban E, Felip E, De Angelis F, et al. Pembrolizumab plus chemotherapy in metastatic non-small-cell lung cancer. *N Engl J Med.* 2018;378(22):2078-92. Doi: 10.1056/NEJMoa1801005.
102. Sledge GW, Toi M, Neven P, Sohn J, Inoue K, Pivot X, et al. The effect of abemaciclib plus fulvestrant on overall survival in hormone receptor-positive, ERBB2-negative breast cancer

- that progressed on endocrine therapy-MONARCH 2: a randomized clinical trial. *JAMA Oncol.* 2019; 6(1):116-124. Doi: 10.1001/jamaoncol.2019.4782.
103. Schoenfeld JD, Hanna GJ, Jo VY, Rawal B, Chen YH, Catalano PS, et al. Neoadjuvant nivolumab or nivolumab plus ipilimumab in untreated oral cavity squamous cell carcinoma: a phase 2 open-label randomized clinical trial. *JAMA Oncol.* 2020;6(10):1563-70. Doi: 10.1001/jamaoncol.2020.2955.
104. Gettinger SN, Redman MW, Bazhenova L, Hirsch FR, Mack PC, Schwartz LH, et al. Nivolumab plus ipilimumab vs nivolumab for previously treated patients with stage IV squamous cell lung cancer: the Lung-MAP S1400I phase 3 randomized clinical trial. *JAMA Oncol.* 2021;7(9):1368-77. Doi: 10.1001/jamaoncol.2021.2209.
105. MedRxiv. Repurposed antiviral drugs for Covid-19 – interim WHO Solidarity trial results [Internet]. 2020 [citado 17 de diciembre de 2022]. Disponible en: <https://www.medrxiv.org/content/10.1101/2020.10.15.20209817v1>
106. Agencia Española de Medicamentos y Productos Sanitarios. Informe de Posicionamiento Terapéutico de atezolizumab (Tecentriq®) en el tratamiento de cáncer de pulmón no microcítico. Agencia Española de Medicamentos y Productos Sanitarios. 2018 [citado 17 de diciembre de 2022]. Disponible en: <https://www.aemps.gob.es/medicamentosUsoHumano/informesPublicos/docs/IPT-atezolizumab-Tecentriq-cancer-pulmon.pdf>
107. Agencia Española de Medicamentos y Productos Sanitarios. Informe de Posicionamiento Terapéutico de pembrolizumab (Keytruda®) en melanoma. Agencia Española de Medicamentos y Productos Sanitarios. 2022 [citado 17 de diciembre de 2022]. Disponible en: https://www.aemps.gob.es/medicamentosUsoHumano/informesPublicos/docs/2022/IPT_15-2022-Keytruda.pdf
108. Cheema PK, Burkes RL. Overall survival should be the primary endpoint in clinical trials for advanced non-small-cell lung cancer. *Curr Oncol.* 2013;20(2):e150-60. doi: 10.3747/co.20.1226.
109. Gil-Sierra MD, Fenix-Caballero S, Sanchez-Hidalgo M, Alarcon de la Lastra Romero C, Briceno-Casado MDP, Rios-Sanchez E, et al. Indirect comparison of pembrolizumab plus

chemotherapy versus pembrolizumab in lung cancer [Internet]. Abstract presentado en el 24º Congreso de la European Association of Hospital Pharmacists (Barcelona). 2019 [citado 17 diciembre 2022]. Disponible en: https://www.eahp.eu/sites/default/files/abstract_book_-_bar19_-_updated_0.pdf

110. Gil-Sierra MD, Fenix-Caballero S, Alegre-del Rey E. Pembrolizumab plus chemotherapy in lung cancer. *N Engl J Med*. 2018;379(11):e18. Doi: 10.1056/NEJMc1808567.
111. Shoushtari AN, Friedman CF, Navid-Azarbaijani P, Postow MA, Callahan MK, Momtaz P, et al. Measuring toxic effects and time to treatment failure for nivolumab plus ipilimumab in melanoma. *JAMA Oncol*. 2018;4(1):98-101. Doi: 10.1001/jamaoncol.2017.2391.
112. Amonkar MM, Chastek B, Samant N, Teitelbaum A. Economic burden of resected squamous cell carcinoma of the head and neck in a US managed-care population. *J Med Econ*. 2011;14(4):421-32. Doi: 10.3111/13696998.2011.584096.
113. Wang DY, Salem JE, Cohen JV, Chandra S, Menzer C, Ye F, et al. Fatal toxic effects associated with immune checkpoint inhibitors: a systematic review and meta-analysis. *JAMA Oncol*. 2018;4(12):1721-8. Doi: 10.1001/jamaoncol.2018.3923.
114. Hu H, She L, Liao M, Shi Y, Yao L, Ding D, et al. Cost-effectiveness analysis of nivolumab plus ipilimumab vs. chemotherapy as first-line therapy in advanced non-small cell lung cancer. *Front Oncol*. 2020;10:1649. Doi: 10.3389/fonc.2020.01649.
115. Fernández-Llamazares CM, López-Briz E. Hospital pharmacist challenges in evaluation of scientific evidence and its incorporation to pharmacotherapeutic protocols through therapeutic committees in COVID-19 times. *Farm Hosp*. 2020;44(7):24-7. Doi: 10.7399/fh.11487.
116. Spanish Group of the European Study on Asthma. Prevalence of bronchial hyperreactivity and asthma in young adults from 5 Spanish areas. *European study of asthma*. *Med Clin (Barc)*. 1996;106(20):761-7.
117. Quirce S, Plaza V, Picado C, Vennera M, Casafont J. Prevalence of uncontrolled severe persistent asthma in pneumology and allergy hospital units in Spain. *J Investig Allergol Clin Immunol*. 2011;21(6):466-71.

118. Comité Ejecutivo de la Guía Española para el Manejo del Asma (GEMA). Guía Española para el Manejo del Asma (GEMA 4.3) [Internet]. Madrid: Luzan. 2018 [citado 12 de diciembre 2019]. Disponible: <https://www.semg.es/index.php/consensos-guias-y-protocolos/279-gema-4-3-guia-espanola-para-el-manejo-delasma>
119. Chung KF, Wenzel SE, Brozek JL, Bush A, Castro M, Sterk PJ, et al. International ERS/ATS guidelines on definition, evaluation and treatment of severe asthma. *Eur Respir J*. 2014;43(2):343-73. Doi: 10.1183/09031936.00202013.
120. Agencia Española de Medicamentos y Productos Sanitarios. Ficha técnica Xolair 150 mg solución inyectable [Internet]. Agencia Española de Medicamentos y Productos Sanitarios. 2005 [citado 17 de diciembre de 2021]. Disponible en: https://cima.aemps.es/cima/dohtml/ft/05319008/FT_05319008.html
121. Agencia Española de Medicamentos y Productos Sanitarios. Ficha técnica Nucala 100 mg polvo para solución inyectable [Internet]. Agencia Española de Medicamentos y Productos Sanitarios. 2015 [citado 17 de diciembre de 2022]. Disponible en: https://cima.aemps.es/cima/pdfs/ft/1151043001/FT_1151043001.pdf
122. European Medicines Agency. Nucala [Internet]. European Medicines Agency. 2018 [citado 17 de diciembre de 2022]. Disponible en: <https://www.ema.europa.eu/en/medicines/human/EPAR/nucala>
123. Tran TN, Khattry DB, Ke X, Ward CK, Gossage D. High blood eosinophil count is associated with more frequent asthma attacks in asthma patients. *Ann Allergy Asthma Immunol*. 2014;113(1):19-24. Doi: 10.1016/j.anai.2014.04.011.
124. Price DB, Rigazio A, Campbell JD, Bleecker ER, Corrigan CJ, Thomas M, et al. Blood eosinophil count and prospective annual asthma disease burden: a UK cohort study. *Lancet Respir Med*. 2015;3(11):849-58. Doi: 10.1016/S2213-2600(15)00367-7.
125. Agencia Española de Medicamentos y productos sanitarios. Informe de Posicionamiento Terapéutico de mepolizumab (Nucala®) como tratamiento adicional en el asma eosinofílica refractaria grave. Agencia Española de Medicamentos y Productos Sanitarios. 2016 [citado 17 de diciembre de 2022]. Disponible en: https://www.aemps.gob.es/medicamentosUsoHumano/informesPublicos/docs/IPT-mepolizumab-Nucala-asma_EPOC.pdf

126. GlaxoSmithKline. Omalizumab to mepolizumab switch study in severe eosinophilic asthma patients [Internet]. US National Library of Medicine. ClinicalTrials.gov. 2016 [acceso 17 de diciembre 2022]. Disponible en: <https://clinicaltrials.gov/ct2/show/NCT02654145>
127. Agencia Española de Medicamentos y Productos Sanitarios. Ficha técnica Cinqaero 10 mg/ml concentrado para solución para perfusión [Internet]. Agencia Española de Medicamentos y Productos Sanitarios. 2016 [citado 17 de diciembre de 2022]. Disponible en: https://cima.aemps.es/cima/pdfs/ft/1161125001/FT_1161125001.pdf
128. Agencia Española de Medicamentos y Productos Sanitarios. Ficha técnica Fasentra 30 mg solución inyectable en jeringa precargada [Internet]. Agencia Española de Medicamentos y Productos Sanitarios. 2018 [citado 17 de diciembre de 2022]. Disponible en: https://cima.aemps.es/cima/pdfs/ft/1171252001/FT_1171252001.pdf
129. López Bastida J, Oliva J, Antoñanzas F, García-Altés A, Gisbert R, Mar J, et al. [A proposed guideline for economic evaluation of health technologies]. *Gac Sanit.* 2010;24(2):154-70. Doi: 10.1016/j.gaceta.2009.07.011.
130. Sullivan SD, Mauskopf JA, Augustovski F, Jaime Caro J, Lee KM, Minchin M, et al. Budget impact analysis-principles of good practice: report of the ISPOR 2012 Budget Impact Analysis Good Practice II Task Force. *Value Health.* 2014;17(1):5-14. Doi: 10.1016/j.jval.2013.08.2291.
131. Instituto Nacional de Estadística. España en cifras 2015 [Internet]. Catálogo de publicaciones oficiales de la Administración General del Estado. 2015 [citado 17 de diciembre 2022]. Disponible en: https://www.ine.es/prodyser/espa_cifras/2015/files/assets/basic-html/page-2.html
132. Consejo General de Colegios Oficiales de Farmaceuticos. BOT Plus 2. Base de Datos de Medicamentos [Internet]. Madrid: Consejo General de Colegios Oficiales de Farmaceuticos. 2013 [citado 17 de diciembre 2022]. Disponible en: <https://botplusweb.portalfarma.com/botplus.aspx>
133. Martínez-Moragón E, Serra-Batlles J, De Diego A, Palop M, Casan P, Rubio-Terrés C, et al. [Economic cost of treating the patient with asthma in Spain: the AsmaCost study]. *Arch Bronconeumol.* 2009;45(10):481-6. Doi: 10.1016/j.arbres.2009.04.006.

134. Moon JC, Flett AS, Godman BB, Grosso AM, Wierzbicki AS. Getting better value from the NHS drug budget. *BMJ*. 2010;341:c6449. Doi: 10.1136/bmj.c6449.
135. Humbert M, Beasley R, Ayres J, Slavin R, Hébert J, Bousquet J, et al. Benefits of omalizumab as add-on therapy in patients with severe persistent asthma who are inadequately controlled despite best available therapy (GINA 2002 step 4 treatment): INNOVATE. *Allergy*. 2005;60(3):309-16. Doi: 10.1111/j.1398-9995.2004.00772.x.
136. Grimaldi-Bensouda L, Zureik M, Aubier M, Humbert M, Levy J, Benichou J, et al. Does omalizumab make a difference to the real-life treatment of asthma exacerbations?: Results from a large cohort of patients with severe uncontrolled asthma. *Chest*. 2013;143(2):398-405. Doi: 10.1378/chest.12-1372.
137. McQueen RB, Sheehan DN, Whittington MD, van Boven JFM, Campbell JD. Cost-effectiveness of biological asthma treatments: a systematic review and recommendations for future economic evaluations. *Pharmacoeconomics*. 2018;36(8):957-71. Doi: 10.1007/s40273-018-0658-x.
138. Cockle SM, Stynes G, Gunsoy NB, Parks D, Alfonso-Cristancho R, Wex J, et al. Comparative effectiveness of mepolizumab and omalizumab in severe asthma: an indirect treatment comparison. *Respir Med*. 2017;123:140-8. Doi: 10.1016/j.rmed.2016.12.009.
139. Nachef Z, Krishnan A, Mashtare T, Zhuang T, Mador MJ. Omalizumab versus mepolizumab as add-on therapy in asthma patients not well controlled on at least an inhaled corticosteroid: a network meta-analysis. *J Asthma*. 2018;55(1):89-100. Doi: 10.1080/02770903.2017.1306548.
140. Canadian Agency for Drugs and Technologies in Health. CDR Clinical Review Report for Nucala® [Internet]. Canadian Agency for Drugs and Technologies in Health. 2016 [citado 18 de diciembre 2022]. Disponible en: https://www.cadth.ca/sites/default/files/cdr/clinical/SR0461_Nucala_CL_Report_e.pdf
141. Bermejo I, Stevenson M, Cooper K, Harnan S, Hamilton J, Clowes M, et al. Mepolizumab for treating severe eosinophilic asthma: an evidence review group perspective of a NICE single technology appraisal. *Pharmacoeconomics*. 2018;36(2):131-44. Doi: 10.1007/s40273-017-0571-8.

142. Whittington MD, McQueen RB, Ollendorf DA, Tice JA, Chapman RH, Pearson SD, et al. Assessing the value of mepolizumab for severe eosinophilic asthma: a cost-effectiveness analysis. *Ann Allergy Asthma Immunol.* 2017;118(2):220-5. Doi: 10.1016/j.anai.2016.10.028.
143. Tice JA, Ollendorf DA, Campbell JD, Chapman R, Shore KK, Weissberg J, et al. Mepolizumab (Nucala®, GlaxoSmithKline plc.) for the treatment of severe asthma with eosinophilia: effectiveness, value, and value-based price benchmarks: final report [Internet]. Institute for Clinical and Economic Review ICER. 2016 [citado 18 de diciembre 2022]. Disponible en: <https://collections.nlm.nih.gov/master/borndig/101678996/Mepolizumab%20Nucala%20for%20the%20Treatment%20of%20Severe%20Asthma%20with%20Eosinophilia.pdf>
144. Gil-Sierra MD, Briceño-Casado MDP, Fénix-Caballero S. Analysis of overall survival benefit of abemaciclib plus fulvestrant in hormone receptor-positive, ERBB2-negative breast cancer. *JAMA Oncol.* 2020;6(7):1122. Doi: 10.1001/jamaoncol.2020.1516.
145. Gil-Sierra MD, Sánchez-Hidalgo M, Alarcón de la Lastra-Romero C. Clinical decision-making about neoadjuvant nivolumab plus ipilimumab. *JAMA Oncology.* 2021;7(2):309-309. Doi: 10.1001/jamaoncol.2020.6989.
146. Gil-Sierra MD, Briceño-Casado MDP, Sánchez-Hidalgo M. Nivolumab plus ipilimumab for previously treated stage IV squamous cell lung cancer. *JAMA Oncol.* 2022;8(4):642. Doi: 10.1001/jamaoncol.2021.7787.
147. Mennini FS, Bini C, Marcellusi A, Del Vecchio M. Cost estimate of immune-related adverse reactions associated with innovative treatments of metastatic melanoma. *Clin Drug Investig.* 2018;38(10):967-76. Doi: 10.1007/s40261-018-0690-9.
148. Agencia Española de Medicamentos y Productos Sanitarios. Ficha técnica Yervoy 5 mg/ml concentrado para solución para perfusión [Internet]. Agencia Española de Medicamentos y Productos Sanitarios. 2011 [citado 18 de diciembre de 2022]. Disponible en: https://cima.aemps.es/cima/pdfs/ft/11698001/FT_11698001.pdf
149. Chi KN, Agarwal N, Bjartell A, Chung BH, Pereira de Santana Gomes AJ, Given R, et al. Apalutamide for metastatic, castration-sensitive prostate cancer. *N Engl J Med.* 2019;381(1):13-24. Doi: 10.1056/NEJMoa1903307

150. Gil-Sierra MD, Briceño-Casado MDP, Sánchez-Hidalgo M. Clinical and pharmacoeconomic impact of subgroup analysis in onco-hematological patients. *Support Care Cancer*. 2022;30(5):3761-72. Doi: 10.1007/s00520-022-06823-y
151. Turner NC, Slamon DJ, Ro J, Bondarenko I, Im SA, Masuda N, et al. Overall survival with palbociclib and fulvestrant in advanced breast cancer. *N Engl J Med*. 2018;379(20):1926-36. Doi: 10.1056/NEJMoa1810527.
152. Slamon DJ, Neven P, Chia S, Fasching PA, De Laurentiis M, Im SA, et al. Phase III randomized study of ribociclib and fulvestrant in hormone receptor-positive, human epidermal growth factor receptor 2-negative advanced breast cancer: MONALEESA-3. *J Clin Oncol*. 2018;36(24):2465-72. Doi: 10.1200/JCO.2018.78.9909
153. García-Mochón L, Gil-Sierra MD, Alegre-Del Rey EJ, Alarcón de la Lastra-Romero C, Sánchez-Hidalgo M. Economic evaluation and budgetary burden of mepolizumab in severe refractory eosinophilic asthma. *Farm Hosp*. 2019;43(6):187-93. Doi: 10.7399/fh.11221.
154. Gil-Sierra MD, Gimeno-Ballester V, Fenix-Caballero S, Alegre-Del Rey EJ. Network meta-analysis of first-line treatments in transplant-ineligible multiple myeloma patients. *Eur J Haematol*. 2020;105(1):56-65. Doi: 10.1111/ejh.13407
155. Gil-Sierra MD, Briceño-Casado MP, Sierra-Sanchez JF. Ivosidenib and azacitidine in IDH1-mutated AML. *N Engl J Med*. 2022;386(26):2535-6. Doi: 10.1056/NEJMc2206489.
156. Gil-Sierra MD, Gimeno-Ballester V, Briceno-Casado MDP, Sanchez-Hidalgo M, Alarcon de la Lastra-Romero C, Fenix-Caballero S, et al. Network metaanalysis of first-line antiangiogenic drugs in advanced renal cell carcinoma [Internet]. Abstract presentado en el 24º Congreso de la European Association of Hospital Pharmacists (Barcelona). 2019 [citado 18 diciembre 2022]. Disponible en: https://www.eahp.eu/sites/default/files/abstract_book_-_bar19_-_updated_0.pdf