**RESUMEN GRADE DE LA REVISIÓN SISTEMÁTICA:**

**Efecto del control intensivo de la glucemia en resultados cardiovasculares y muertes en pacientes con DM: Metaanálisis de ensayos clínicos.**

Ray K et al. Effect of intensive control of glucose on cardiovascular outcomes and death in patients with diabetes mellitus: a meta-analysis of randomised controlled trials. Lancet. 2009;373:1765–72.

**Abreviaturas:** **AAR**: aumento absoluto del riesgo; **ACV**: accidente cerebrovascular; **CV**: cardiovascular; **DE**: desviación estándar; **ECA:** estudio aleatorizado y controlado; **EE**:error estándar; **IAM**: infarto agudo de miocardio; **IC**: intervalo de confianza; **ITT:** análisis de los resultados según el criterio de intención de tratar: **Mort**: mortalidad (por todas las causas); **NND**: número necesario a tratar para dañar a 1 paciente; **NNT**: número necesario a tratar para evitar 1 evento; OR: odds ratio; **RAR**: reducción absoluta del riesgo; **RR**: riesgo relativo; **RRR**: reducción relativa del riesgo.

**NOTA:** Los intervalos los calculamos para un 95% de confianza (IC 95%), salvo que indiquemos expresamente otro porcentaje. Calculadora para los estudios clínicos individuales disponible en la web evalmed.es, pestaña HERRAMIENTAS.

***I. INTRODUCCIÓN.***

Según informan los autores, la disminución de eventos macrovasculares en pacientes DM2 sometidos a control intensivo de glucemia no ha podido ser demostrada por los ECA individuales, lo cual podría ser debido a que sus tamaños muestrales no alcanzan la potencia suficiente para establecer la diferencia de efectos respecto al control convencional. Un metaanálisis puede aclarar esa duda.

***II. LO PROYECTADO.***

**A) REFERENCIA (AUTOR, PUBLICACIÓN).**

Ray K et al. Lancet. 2009;373:1765–72.

**B) OBJETIVO:** Investigar, mediante una revisión sistemática de varios ECAs (con la intención de utilizar la técnica de metaanálisis), si, en pacientes con DM2, el tratamiento intensivo de la glucemia frente al tratamiento convencional, se reducen: **1)** los eventos macrovasculares [a) IAM no fatal; b) enfermedad coronaria (IAM fatal y no fatal); y c) ACV fatal y no fatal; y **2)** la [Mortalidad por todas las causas].

**Duración de los estudios:** Los investigadores deberían haberse planteado la pregunta para un margen en el tiempo de seguimiento, pues sin el tiempo en realidad la pregunta no está terminada, no es clara ni precisa (ver en nuestra tabla PICO la diferencia entre los tiempos de seguimiento de los estudios evaluados y los porcentajes de eventos /año).

**C) TIPO DE ESTUDIO:**

Antes de continuar, debemos saber si hay **heterogeneidad clínica** entre los estudios. Para estimarla se elabora una **tabla PICO orientada a GRADE**, que contiene las características basales de todos los pacientes, la validez de los estudios individuales (GRADE), la homogeneidad en las características sociodemográficas y clínicas en el inicio (baseline), los % eventos/ año en los grupos de control, el efecto Hawthorne en el seguimiento, la definición de los eventos, y el rigor para la asignación tales eventos (Comités de Asignación de Eventos).

Revisión sistemática con 5 ECA, con **metaanálisis de efectos aleatorios**. Medida de asociación OR (aunque nosotros lo presentamos en RR por ser más fácil de interpretar).

El **modelo de efectos fijos** considera que el efecto “RR” es igual en todos los estudios, y que la única desviación que haya en cada ECAi se debe únicamente a su varianza dentro, de modo que se cada efecto particular de cada ECAi se representa así:

RRi = “RR” + error-intrai

Como consecuencia de la combinación de varios efectos RRi de los ECAi de un metaanálisis de efectos fijos, el efecto combinado obtenido **RRobtenido-por-MA** tiene un intervalo de confianza que se debe únicamente a la oportuna combinación de las varianzas intra de cada uno de los ECAi, refiriéndonos a que en cada ensayo clínico ECAi tiene su particular varianza, propia de la distribución de los rasgos de los pacientes respecto a la media de cada uno de ellos (“error-intraestudioi”).

El **modelo de efectos aleatorios** considera que los efectos RRi de los ECAi son muestras tomadas de una Población de Efectos (y que no hay un efecto fijo). Los efectos RRi de los ECAi siguen una distribución normal alrededor del Efecto Promedio “Promedio-de-RRi”, en la que hay una varianza entre los particulares efectos RRi de los ECAi distribuidos, que denominamos error-interefectos. Como dentro de cada estudio ya hemos mencionado que previamente hay varianza una intra estudio, cada efecto particular de cada ECAi se representa así:

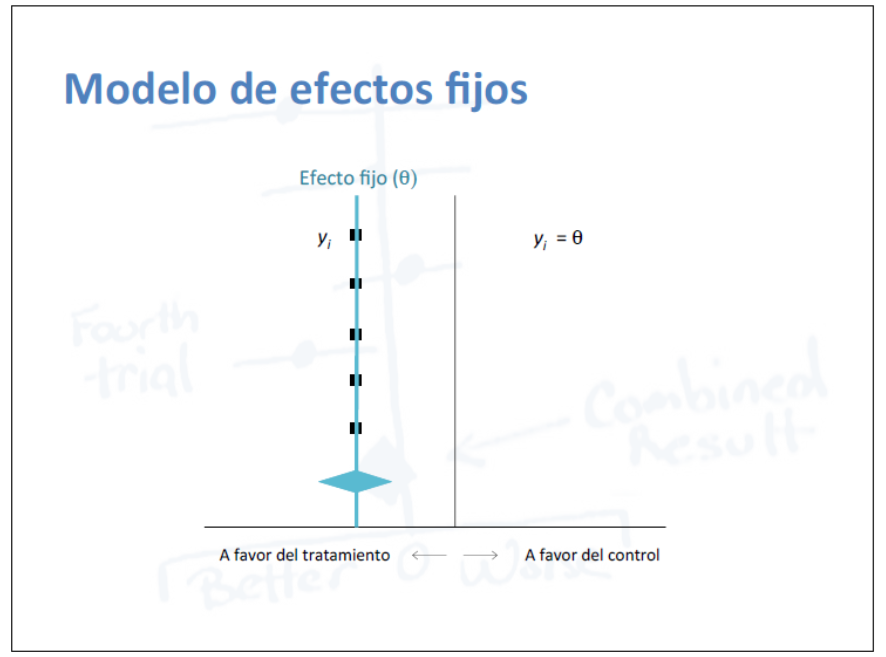
RRi = “Promedio-de-RRi” + error-intraestudioi + error-interefectos

Como consecuencia de la combinación de varios efectos RRi de los ECAi de un metaanálisis de efectos aleatorios, el efecto combinado obtenido **RRobtenido-por-MA** tiene un intervalo de confianza que se debe a dos varianzas: a) la que surge de la combinación de los errores intra estudio; y b) la que surge de la varianza inter-efectos.

Se entiende por tanto que, a igual valor de la estimación puntual del **RRobtenido-por-MA**, el intervalo de confianza de efectos aleatorios es más amplio (más conservador) que el de efectos fijos.

Cuando el **índice *I2* de heterogeneidad estadística** es bajo, la estimación puntual del **RRobtenido-por-MA** en el modelo de efectos fijos es muy similar a la estimación puntual en el modelo de efectos aleatorios. Cuando el **índice *I2*** es nulo, significa que el error-interefectos = 0, y entonces ambos modelos coinciden en la estimación puntual y en los intervalos de confianza.

Conviene hacer los cálculos con los dos modelos para ver si cambia el sentido del efecto. Ray y col estimaron que los efectos entre los estudios podrían no ser fijos, por lo que hicieron sus cálculos por **el modelo de efectos aleatorios, más conservador, pero con menos riesgos de sobreestimación del efecto**.







La **heterogeneidad estadística** entre ensayos informó el **índice de heterogeneidad *I²***.

Este índice de heterogeneidad *I²* se obtiene a través de la prueba “Q de Cochran”. La Q tiene una distribución como una χ² con k-1 grados de libertad, siendo k el número de estudios. Por ello el valor de *p* se extrae con excell igual que la *p* de χ² [*p*=DISTR.CHI(“Q”;grados libertad)]; si p < 0,05 => Heterogeneidad; sí p > 0,05 => Homogeneidad. Suele, no obstante, utilizarse en estos casos una *p* de 0,1 en lugar de 0,05 cuando son pocos estudios.

Se calcula así: Q = suma {wi \* [ln(RRi) – ln(RRMH )]2}, siendo wi el peso de cada uno de los estudios (wi = 1/si2, es decir el inverso de la varianza, según Cochran 1937, y wi = (bi x ci)/ni, según Mantel-Haenszel) y el RRMH = suma(wi x RRi) / suma(wi).

El problema con la Q es que presenta poca potencia estadística cuando los estudios son pocos. Para solventarlo, se inventó el **índice de heterogeneidad *I2***, cuyo cálculo es muy sencillo. Si Q > k-1, entonces *I2* = [Q-(k-1)] / Q \* 100. Si Q ≤ -1, entonces *I2* = 0. Mediante otros cálculos se obtienen los IC 95% de la *I2*.

El estadístico *I2* indica la proporción en la estimación del efecto del tratamiento que es debido a la heterogeneidad entre estudios. Una *I²* del 23% indica el 23% de la estimación del efecto es debida a la heterogeneidad entre estudios (y no al error de muestreo aleatorio). El mejor resultado es el 0%, y Higgins y Thompson proponen interpretar los porcentajes 25%, 50% y 75% como heterogeneidad muy baja, baja, moderada y alta, respectivamente.

Matemáticamente es así: *I2* = *2* / (*2* + s*2*), donde *2* corresponde a la varianza entre estudios (inter-estudios), es decir la varianza observada en los RR de los estudios respecto al **RRobtenido-por-MA**,, y *s2* a la varianza obtenida de todas las varianzas intra-estudio, siendo cada varianza interna de cada estudio la que es debida fundamentalmente a la variabilidad entre pacientes.

La **probabilidad de sesgo de publicación** se evaluó con el diagrama de embudo (funnel plot) y el Egger test (que da un valor de *p*).



**D) ESTRATEGIA DE BÚSQUEDA, SELECCIÓN E INCLUSIÓN DE LOS ESTUDIOS QUE DEFINEN CLARAMENTE LOS PACIENTES, LA EXPOSICIÓN, LA COMPARACIÓN Y LOS RESULTADOS.**

**1º Elegibles:** Se buscó en Medline, Cochrane Central y EmBase, por artículos **en inglés** desde enero-1970 a enero-2009, con términos relacionados con diabetes y resultados vasculares (ejemplo, “enfermedades cardiovasculares”, “diabetes mellitus”, “glucosa” y “Hb1Ac”). Se restringió a ECAs, lo que proporcionó 2.439 artículos desde: a) títulos; b) abstracts; c) texto completo; o combinación de a, b y c. Se complementó con una búsqueda electrónica de listados de artículos relevantes, incluyendo metaanálisis, revisiones y discusiones con expertos.

**2º Criterios de inclusión:**

1) Asignación aleatorizada de los pacientes con DM2 a control intensivo de glucemia frente a convencional, con significativa diferencia de control glucémico (medido por Hb1Ac) entre los grupos de pacientes durante el seguimiento.

2) Resultados medidos con la variable primaria, basada en eventos cardiovasculares, y completa información acerca de los efectos estimados o provisión de información para permitir el cálculo de la estimación del efecto de: a) IAM no fatal; b) enfermedad coronaria (IAM fatal y no fatal); c) ACV fatal y no fatal; y d) mortalidad por cualquier causa.

3) Que sean sólo de individuos estables, excluyendo estudios en salas de hospitalización de agudos.

Según los autores, cumplieron los criterios de inclusión 16 artículos de 11 ECA.

**3º Exclusiones de estudios y motivos:** Seis de los once ECA se excluyeron: **a)** el ADOPT y el RECORD no evaluaron resultados cardiovasculares en la variable primaria, y RECORD sólo tuvo datos intermedios para algunos de los resultados sin proveer los cambios de Hb1Ac durante el seguimiento; **b)** el DRAM fue llevado a cabo en individuos con la tolerancia a la glucosa alterada; **c)** el UGDP incluyó pacientes con diabetes y tolerancia a la glucosa alterada, y además no proporcionó información separada del grupo de diabetes ni los efectos estudiados de los resultados de interés en cada grupo; **d)** el STENO 2 analizó varias intervenciones y por tanto no evaluó el control intensivo de la glucemia con el convencional; además informó una variable principal compuesta de eventos cardiovasculares, incluyendo enfermedad vascular periférica y angina, en lugar de variables individuales de interés (además los eventos incluidos en la variable primaria de este estudio no estaban claramente adjudicados).

**E) ASIGNACIÓN DE LOS PACIENTES A LOS GRUPOS EN LOS ESTUDIOS EVALUADOS.**

**1º ¿Aleatorización de los estudios evaluados?:** Todos eran estudios aleatorizados.

**2º ¿Se mantuvo oculta la secuencia de aleatorización para pacientes y los médicos que hacen el seguimiento en los estudios evaluados?:** Únicamente se mantuvo en el PROactive. En los otros no (UKPDS, ADVANCE, VADT y ACCORD) porque eran abiertos.

**3º ¿Fueron similares los grupos de intervención y control con respecto a los factores pronósticos conocidos de los estudios evaluados?:** Lo primero que debe evaluarse es la primera parte de la **Tabla 1 (Esquema PICO)**, concretamente las características en el inicio.

Al observarla, vemos diferencias (tiempo desde el diagnóstico, % varones, % enfermedad CV) que pueden afectar a la generalización del resultado del metaanálisis (posible heterogeneidad clínica).

Esto no quiere decir que el metaanálisis no deba comenzarse, porque caben análisis de subgrupos (por ejemplo, de nuevo diagnóstico y de antiguo diagnóstico) que salven la heterogeneidad clínica.

**F) GRUPO DE INTERVENCIÓN Y DE CONTROL DE LA REVISIÓN SISTEMÁTICA PROCEDENTES DE LOS ESTUDIOS SELECCIONADOS.**

**1º Grupo A de intervención de la Revisión Sistemática:** 17.267 pacientes control intensivo de la glucemia.

**2º Grupo B de control de la Revisión Sistemática**: 15.773 pacientes control convencional de la glucosa.

**Comentarios:**

1) Estimamos que el PROactive no cumple los criterios de inclusión porque no estudia control intensivo de la glucosa (perseguir una Hb1Ac <6,5%) frente a control convencional, pues el target de Hb1Ac en ambos era el mismo, ya que la única diferencia que se pretendió fue el efecto de la pioglitazona y el de placebo, y en ambos brazos los investigadores fueron alentados para perseguir una Hb1Ac <6,5%.

2) El UKPDS utilizó 411 pacientes del grupo de control dos veces. El UKPDS tiene para todo el grupo de control 1.138 pacientes. En el UKPDS-33, los **2.729 pacientes** del grupo de intervención se enfrentan a los **1.138 pacientes** del grupo de control. En el UKPDS-34A, los **342 pacientes obesos** del grupo de intervención se enfrentan a **411 pacientes obesos del grupo de control**, pero estos 411 son un subgrupo de obesos de los 1.138 pacientes del grupo de control anterior. Por esa razón, hubo 4.209 individuos totales, distribuidos en 3.071 individuos en el grupo de control intensivo (2.729 del UKPDS y 334 del UKPDS 34A) y 1.138 en el grupo de control convencional. El metaanálisis debe modificarse de modo que no se dupliquen las incidencias de los 411 individuos duplicados por Ray.

3) Si a los 17.267 / 15.773 de Ray restamos 2.505 / 2633 del PROactive y los 0 / 411 del UKPDS, los evaluables serían 14.662 / 12.729 pacientes.

**G) PAUTA DE TRATAMIENTOS DE LOS ESTUDIOS EVALUADOS.**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **UKPDS** | **PROactive** | **ADVANCE** | **VADT** | **ACCORD** |
| **Intensive** | Sulphonylurea, insulin, or metformin. Target FPG <108 mg/dl | 15–45 mg oral pioglitazone plus current medication | 30–120 mg oral gliclazide modified release, with or without metformin, thiazolidinedione, glinide, acarbose, or insulin. Target HbA1c concentration ≤6,5% | Maximum dose of metformin, with either rosiglitazone (BMI >27) or glimepiride and rosiglitazone (BMI <27). | Treatment with metformin, sulphonylurea, glinide, thiazolidinedione, acarbose, insulin, or a combination of these. Target HbA1c concentration <6% |
| **Standard** | Standard diet. Target FPG <270 mg/dl | Current medication | Standard treatment as per local guideline | Half-dose of intensive treatments. | Standard treatment. Target HbA1c concentration 7–7,9% |

**H) VARIABLES DE MEDIDAS DE RESULTADO BUSCADAS EN LA REVISIÓN SISTEMÁTICA.**

**1º Variables primarias de la Revisión Sistemática:** 1) IAM no fatal; 2) enfermedad coronaria (IAM fatal y no fatal); 3) ACV fatal y no fatal; y 4) mortalidad por todas las causas.

**2º Variables secundarias de la Revisión Sistemática:** 1) Efectos adversos: a) hipoglucemias; b) hipoglucemias graves; y c) cambio de peso corporal.

**Comentarios:** Tiene contabilizados todos los “IAM fatales” para construir la variable “Enfermedad coronaria” y sin embargo no los toma como variable de interés. Ante esta anomalía, nosotros hemos introducido esta variable para estimar si las dos variables individuales que forman la variable “Enfermedad coronaria” son significativas.

***III. LO CONSEGUIDO.***

**A) RESULTADOS.**

**1º En los estudios evaluados.**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **UKPDS** | **PROactive** | **ADVANCE** | **VADT** | **ACCORD** |
| **¿****Cegado para los que asignan los eventos?** | Sí | Sí | Sí | Sí | Sí |
| **Abandonos y pérdidas.** | 4,40% vs 3,9% | 16,4% vs 16,7% | 0,13% vs 0,18% | 6,5% vs 5,3% | 2,1% vs 2,0% |
| **Análisis por** | ITT | ITT | ITT | ITT | ITT |

**2º Los resultados extraídos de los estudios evaluados, ¿fueron recolectados y asignados por dos o más investigadores de la Revisión Sistemática independientemente? ¿Cómo resolvieron las discrepancias? ¿Los estudios originales estaban cegados para los investigadores de la Revisión Sistemática?:** Se recolectaron por dos investigadores y las discrepancias las adjudicó un tercero.

El artículo no cita si el grado de concordancia entre investigadores se midió con el Coeficiente de concordancia kappa de Cohen. El artículo tampoco indica si las revistas donde se publicaron los estudios originales estaban cegadas para los investigadores de la Revisión Sistemática.

**4º Magnitud y precisión de los resultados de las variables de la Revisión Sistemática:** En las **tablas 3 a 8** los evaluadores mostramos todos los resultados con nitidez.

1. Beneficios (riesgos evitados):

1) Mortalidad por todas las causas, RR 1,01 (0,89-1,13), diferencia no significativa.

2) ACV totales, RR 0,95 (0,85-1,07), diferencia no significativa

3) IAM fatal: RR 0,88 (0,77-1,01), diferencia no significativa

4) IAM no fatal, con una heterogeneidad *I2*= 0%, se obtuvo un RR 0,84 (0,76-0,93), que aplicado al 1,00% de “eventos / paciente-año” estimado para un tiempo global de seguimiento de 5,01 años, se nos convierte en un NNT 625 (417 a 1429) por año (cuya magnitud de efecto estimamos muy baja), que podría expresarse en un NNT 125 (83 a 285) en 5 años.

5) La variable “Enfermedad coronaria (IAM fatal + no fatal)”, con una heterogeneidad *I2*=0%, muestra un beneficio estadísticamente significativo, RR 0,85 (0,79-0,92), pero no la describimos porque es la suma de las dos anteriores, y el beneficio que muestra parece ser debido a los IAM no fatales, porque los IAM fatales son no significativos.

2. Riesgos añadidos

1) Hipoglucemia grave (que requiere asistencia médica): Con una heterogeneidad *I2*=76% (muy alta) obtenemos un RR 2,23 (1,72-2,89), con un NND 149 (97 a 255) por año; que puede expresarse como NND 30 (19 a 51) en 5 años.

Sin embargo el *I2*=76% nos indica que la considerable heterogeneidad entre los estudios es muy improbable que sea debida al azar.

En este caso, se considera más adecuado detallar cada uno de los estudios individualmente: a) ACCORD 2008, RR 3,00 (2,55-3,54, NND 14 (13 a 17) en 3,8 años; b) ADVANCE 2008, RR 1,85 (1,42-2,42), NND 81 (57 a 143) en 5 años; c) PROactive 205, RR 1,75 (0,83-3,66), NNT -321 (-137 a 739) en 2,8 años; d) UKPDS 1998, RR 1,90 (1,56-2,33), NND 15 (12 a 21) en 10 años; y e) VADT 2009, RR 2,59 (1,69-3,97), NND 20 (14 a 36) en 5,6 años.

2) Hipoglucemias (que requieren cualquier forma de asistencia): 38,1% vs 28,6%

3) Cambio en el peso corporal: +2,5 Kg respecto al convencional.

**Comentarios:** Es sabido que el número de eventos sigue una distribución que no es lineal y que el número suele ser más denso cuanto mayor sea el tiempo se seguimiento. Sin embargo, los investigadores mezclan el número de eventos acumulados en distintos tiempos de seguimiento, como si la densidad de eventos hubiera sido constante a lo largo del tiempo: a) UKPDS 33 A: 10 años; b) UKPDS 34 A: 6,6 años; c) PROactive: 2,9 años; d) ADVANCE: 5 años; d) VADT: 5,6 años; y e) ACCORD: 3,8 años.

**6º ¿Se hizo análisis de sensibilidad por los investigadores de la Revisión Sistemática?:** Los autores dicen que sí, refiriéndose a que compararon los OR con los RR, viendo que las magnitudes eran similares, lo cual es una tautología porque es el mismo número expresado en dos marcos distintos. Un análisis de sensibilidad es mucho más que esto, y nosotros particularmente practicamos varios, como explicamos en otro apartado.

**7º ¿Se hizo análisis de sensibilidad por los que hacen la lectura crítica (grupo evalmed-GRADE):** Los beneficios de la variable “IAM no fatal” (así como de “Enfermedad coronaria”) siguen siendo estadísticamente significativos eliminando de la revisión primero el UKPDS, y a continuación también el PROactive.

**B) FINANCIACIÓN DE LA REVISIÓN Y CONFLICTOS DE INTERESES:** Ninguna entidad financió los fondos para la realización de este estudio.

Conflictos de intereses: Ray declara haber recibido honorarios de Novartis por conferencias y actuaciones como miembro de comités asesores. Sattar declara haber recibido honorarios de Merck, GSK, MSD y Novo Nordisks. Los demás declaran no tener conflictos de intereses.

Ray, en opinión de Montaño (que ha hecho una búsqueda del perfil comunicativo de Ray), tiene un conflicto de interés “de posición”, porque ya era un firme defensor de la medicalización intensiva del diabético.

***IV. COMENTARIOS (DISCUSIÓN Y OPINIÓN DEL EVALUADOR).***

**1º ¿Fueron los pacientes del estudio similares a los que yo atiendo?**

Las características en el inicio son heterogéneas (hay pacientes DM2 desde recién diagnosticados a 9 años de enfermedad). No sé a qué pacientes míos se puede aplicar, porque no se ha hecho análisis de subgrupos DM2 nuevos y DM2 antiguos, o por la edad, o por los factores pronósticos, basados en la edad y otros factores de riesgo cardiovascular.

Los investigadores, entre las limitaciones del estudio, citan que no hay datos crudos de pacientes individuales disponibles para analizar los efectos del control glucémico intensivo en varios subgrupos de pacientes (por ejemplo, por edad, hombres frente mujeres, duración de la diabetes, Hb1Ac de partida, prevalencia de enfermedad cardiovascular en el inicio, comorbilidad). Estos subgrupos nos hubieran podido ayudar a ver si su perfil es aplicable a “mis pacientes”.

Es difícil pasar a la decisión compartida con “mis pacientes”, incorporando los valores y preferencias, porque sin análisis de subgrupos, no hay un buen firme del que partir.

Richard Lehman dice que este metaanálisis viene a oscurecer lo que ya sabíamos. Este metaanálasis mezcla viejos con nuevos DM2 (y además no lo resuelve mediante un análisis de subgrupos).

Este metaanálisis tiene heterogeneidad CLÍNICA.

**2º ¿Se consideraron todos los resultados de importancia para los pacientes?:**  Sí les importan, porque contiene los efectos beneficiosos, pero también contiene los efectos adversos que les interesan a mis pacientes.

**3º ¿Son creíbles los resultados de subgrupos postulados?:** Sólo hay un análisis de subgrupos: subgrupo de pioglitazona (ACCORD y PROactive) frente a subgrupo sin pioglitazona (UKPDS, ADVANCE y VADT) que muestra una asociación significativa entre insuficiencia cardíaca y pioglitazona.

**4º ¿Cuál es la calidad global de la evidencia para los resultados clave que estoy estudiando?:**

1) La validez de la evidencia GRADE de los estudios originales era ésta:



Y hemos mostrado la validez GRADE combinada de cada una de las variables, que resultó ser en las todas las analizadas “Alta-Moderada”, como puede observarse en las **tablas 3 a 8** de resultados.

2) Al interaccionar la validez GRADE “Alta-Moderada” de las variables analizadas los siguientes motivos de rebaja de la validez estimamos una **validez Moderada-Baja**. Tales motivos son: a) utilizar dos veces los 411 pacientes el control UKPDS 34 A; b) no haber debido incluir el PROactive porque no cumple un criterio de inclusión; c) heterogenedad clínica por mezclar reciente diagnóstico de DM2 con diagnóstico hace 12 años; d) la no inclusión de resultados en salud de interés que están a su alcance (IAM fatal); y e) no haber contabilizado los intervalos de predicción, pues en el IAM no fatal es no estadísticamente significativo

**5º ¿Justifican los beneficios que se esperan del tratamiento los riesgos potenciales, los inconvenientes y los costes del mismo?:** Estimamos que no para una mayoría de los pacientes.















