

El todo es mayor que la suma de las partes

Esta es otra de esas frases famosas que están hasta en la sopa. Al parecer, el primero que tuvo esta ocurrente idea fue Aristóteles, que resumió con ella el principio general del holismo en sus escritos sobre metafísica. ¿Quién hubiese dicho que la frasecita encierra tanta sabiduría? Y es que el holismo insiste en que el todo debe ser considerado de una forma global, ya que sus componentes pueden actuar de una forma sinérgica, permitiendo que el conjunto tenga más significado que el aportado por cada parte de forma individual.

No temáis, seguid en el blog del seso y no en uno de filosofía. Y tampoco he cambiado la temática del blog, pero es que este principio me viene al pelo para introducir esa fiera corrupta del método científico que es la revisión sistemática, con o sin metanálisis.

Vivimos en la era de la información. Desde finales del pasado siglo XX hemos sido testigos de una verdadera explosión de las fuentes de información disponibles, accesibles desde múltiples fuentes y plataformas. El resultado final es que nos vemos desbordados cada vez que necesitamos una información sobre un punto concreto, no sabemos dónde buscar ni cómo podemos encontrar lo que queremos. Por este motivo empezaron a desarrollarse sistemas que sintetizasen la información disponible para hacerla más abarcable cuando se necesitase.

Nacen así las primeras revisiones, llamadas [revisiones narrativas](#) o de autor. Para elaborarlas, uno o varios autores, habitualmente expertos en un tema concreto, hacían una revisión general sobre dicho tema, aunque sin ningún criterio estricto sobre la estrategia de búsqueda ni de selección de la información. Siguiendo con total libertad, los autores analizaban los resultados tal como les mandaba su entendimiento y terminaban sacando sus conclusiones a partir de una síntesis cualitativa de los resultados obtenidos.

Estas revisiones narrativas son muy útiles para adquirir una visión general del tema, especialmente cuando uno sabe poco sobre la materia, pero son poco útiles para el que ya conoce el tema y necesita respuesta a una pregunta más concreta. Además, como todo el procedimiento se hace según el gusto de los autores, las conclusiones son poco reproducibles.

Por estos motivos, una serie de mentes privilegiadas inventan el otro tipo de revisión en el que nos vamos a centrar en esta entrada: la [revisión sistemática](#). En lugar de revisar un tema general, las revisiones

sistemáticas sí se centran en un tema concreto para poder solucionar dudas concretas de la práctica clínica. Además, emplean una estrategia de búsqueda claramente especificada y unos criterios de inclusión de trabajos explícitos y rigurosos, lo que las hace altamente reproducibles si a otro grupo de autores se les ocurre repetir la revisión sobre el mismo tema. Y, por si fuera poco, siempre que es posible van más allá del análisis de síntesis cualitativa, completándola con una síntesis cuantitativa que recibe el simpático nombre de [metanálisis](#).

La elaboración de una revisión sistemática consta de seis pasos: formulación del problema o pregunta que quiere responderse, búsqueda y selección de los trabajos existentes, evaluación de la calidad de estos estudios, extracción de los datos, análisis de los resultados y, finalmente, interpretación y conclusión. Vamos a detallar un poco todo este proceso.

Toda revisión sistemática que se precie debe tratar de responder a una pregunta concreta que debe ser relevante desde el punto de vista clínico. La pregunta se hará habitualmente de forma estructurada con los componentes habituales de población, intervención, comparación y resultado (PICO), de forma que el análisis de estos componentes nos permitirá saber si la revisión es de nuestro interés.

Además, los componentes de la [pregunta clínica estructurada](#) nos ayudarán a realizar la búsqueda de los trabajos relevantes que existan sobre el tema. Esta búsqueda debe ser global y no sesgada, por lo que evitaremos los posibles sesgos de procedencia excluyendo fuentes por su idioma, revista, etc. Lo habitual es utilizar un mínimo de dos bases de datos electrónicas importantes de uso general, como Pubmed, Embase o la de la Cochrane, junto con las específicas del tema que se esté tratando. Es importante que esta búsqueda se complemente con una búsqueda manual en registros no electrónicos y consultando las referencias bibliográficas de los trabajos encontrados, además de otras fuentes de la llamada literatura gris, como tesis, y documentos de congresos, además de documentos de agencias financiadoras, registros e, incluso, establecer contacto con otros investigadores para saber si existen trabajos aún no publicados.

Es muy importante que esta estrategia se especifique claramente en el apartado de métodos de la revisión, de forma que cualquiera pueda reproducirla con posterioridad, si se le antoja. Además, habrá que especificar claramente los criterios de inclusión y exclusión de los estudios primarios de la revisión, el tipo de diseño que se busca y sus componentes principales (otra vez en referencia al PICO, los componentes de la pregunta clínica estructurada).

El tercer paso es la evaluación de la calidad de los estudios encontrados, que debe hacerse por un mínimo de dos personas de forma

independiente, ayudándose de un tercero (que seguramente será el jefe) para desempatar en los casos en que no haya consenso entre los extractores. Para esta tarea suelen utilizarse herramientas o listas de verificación diseñadas para tal efecto, siendo una de las más utilizadas la herramienta de control de sesgos de la Colaboración Cochrane. Esta herramienta valora cinco criterios de los estudios primarios para determinar su riesgo de sesgo: secuencia de aleatorización adecuada (previene el sesgo de selección), enmascaramiento adecuado (previene los sesgos de realización y detección, ambos sesgos de información), ocultamiento de la asignación (previene el sesgo de selección), las pérdidas durante el seguimiento (previene el sesgo de desgaste) y la información de datos selectiva (previene el sesgo de información). Los estudios se clasifican como de alto, bajo o indeterminado riesgo de sesgo. Es frecuente utilizar los colores del semáforo, marcando en verde los estudios con bajo riesgo de sesgo, en rojo los que tienen alto riesgo de sesgo y en amarillo los que se quedan en tierra de nadie. Cuanto más verde veamos, mejor será la calidad de los estudios primarios de la revisión.

Para la extracción de los datos suelen diseñarse también formularios al uso que suelen recoger datos como fecha, ámbito del estudio, tipo de diseño, etc, además de los componentes de la pregunta clínica estructurada. Como en el caso del paso anterior, conviene que esto se haga por más de una persona, estableciendo el método para llegar a un acuerdo en los casos en que no haya consenso entre los revisores.

Y aquí entramos en la parte más interesante de la revisión, el análisis de los resultados. El papel fundamental de los autores será explicar las diferencias que existan entre los estudios primarios que no sean debidas al azar, prestando especial atención a las variaciones en el diseño, población de estudio, exposición o intervención y resultados medidos. Siempre se podrá hacer un análisis de síntesis cualitativa, aunque la verdadera magia de la revisión sistemática es que, cuando las características de los estudios primarios lo permiten, puede realizarse también una síntesis cuantitativa, llamada metanálisis.

Un metanálisis es un análisis estadístico que combina los resultados de varios estudios independientes pero que tratan de responder a una misma pregunta. Aunque el metanálisis puede considerarse como un trabajo de investigación por derecho propio, lo habitual es que sea parte de una revisión sistemática.

Los estudios primarios pueden combinarse empleando una metodología estadística desarrollada para tal fin, lo que tiene una serie de ventajas. La primera, al combinar todos los resultados de los estudios primarios puede obtenerse una visión global más completa (ya sabéis, el todo es mayor...). La segunda, al combinar aumentamos el tamaño de la muestra, lo que

aumenta la potencia del estudio en comparación con la de los estudios individuales, mejorando la estimación del efecto que queremos medir. En tercer lugar, al extraerse las conclusiones de un número mayor de estudios aumenta su validez externa, ya que al haber implicadas poblaciones diferentes es más fácil generalizar los resultados. Por último, puede permitirnos resolver controversias entre las conclusiones de los diferentes estudios primarios de la revisión e, incluso, contestar a preguntas que no se habían planteado en esos estudios.

Una vez hecho el metanálisis habrá que hacer una síntesis final que integre los resultados de las síntesis cualitativa y cuantitativa con el objetivo de dar respuesta a la pregunta que motivó la revisión sistemática o, cuando esto no sea posible, plantear los estudios adicionales que deben realizarse para poder contestarla.

Pero para que un metanálisis merezca todos nuestros respetos debe cumplir una serie de requisitos que son, básicamente, los que ya le hemos exigido a la revisión sistemática de la que forma parte: el metanálisis debe tratar de contestar una pregunta concreta y debe basarse en toda la información relevante disponible, sin sesgo de recuperación ni sesgo de publicación. Además, deben valorarse los estudios primarios para asegurarnos de que tienen la calidad suficiente y que son lo suficientemente homogéneos como para poder combinarlos. Por supuesto, los datos deben analizarse y presentarse de la forma apropiada. Y, por último, debe tener sentido el que queramos combinar los resultados. El que podamos hacerlo no siempre significa que tengamos que hacerlo si el contexto clínico no lo hace necesario.

¿Y cómo se combinan los estudios?, preguntaráis algunos. Pues esa es la madre del cordero del metanálisis (o una de las madres, que tiene varias), porque hay varias formas posibles de hacerlo.

A cualquiera se le ocurre que la forma más sencilla sería tipo Festival de Eurovisión. Contabilizamos los estudios primarios en los que el efecto positivo obtuvo significación estadística y, si son mayoría, decimos que hay consenso a favor del resultado positivo. Este enfoque es bastante sencillo pero, no me lo negaréis, también bastante chapucero y se me ocurren una serie de inconvenientes a la hora de usarlo. Por un lado, implica que no significativo es sinónimo de falta de efecto, lo cual no siempre tiene porqué ser verdad. Además, no tiene en cuenta la dirección y fuerza del efecto en cada estudio, ni la precisión de los estimadores empleados ni la calidad o las peculiaridades de diseño de cada estudio primario. Así que este tipo de abordaje no parece muy recomendable, aunque nadie nos va a poner una multa si lo usamos de forma informal como primera aproximación antes de decidir cuál es la mejor forma de combinar los resultados de los estudios.

Otra posibilidad es utilizar una especie de prueba de los signos, similar a la de algunas técnicas de estadística no paramétrica. Se cuentan los positivos a favor del efecto, les restamos los negativos y obtenemos nuestra conclusión. La verdad es que este método también parece demasiado simple. No tiene en cuenta los estudios sin significación estadística ni la precisión de los estimadores. Así que tampoco utilizaremos mucho este tipo de abordaje, a no ser que solo sepamos la dirección del efecto de los estudios primarios. También podríamos utilizarlo cuando los estudios primarios son muy heterogéneos para obtener una aproximación, aunque yo no me fiaría mucho de los resultados.

El tercer método es combinar las “pes” (nuestras amadas y sacrosantas p). Esto podría pasársenos por la cabeza si tuviésemos una revisión sistemática cuyos estudios primarios usaran diferentes medidas de resultado, aunque todos tratasen de contestar a una misma pregunta. Pensad, por ejemplo, un estudio sobre osteoporosis donde unos miden densitometría ultrasónica, otros DEXA en columna, otros en fémur, etc. El problema de este método es que no tiene en cuenta la intensidad del efecto, sino solo su dirección y su nivel de significación estadística, y todos conocemos las deficiencias de nuestras santas “pes”. Para utilizarlo tendremos que utilizar programas informáticos que combinarán los datos siguiendo una distribución de ji-cuadrado o una normal, dándonos el estimador global con su intervalo de confianza.

El cuarto y último método, que yo me sé, es también el más elegante: hacer una combinación ponderada del efecto estimado en los diferentes estudios primarios. La forma más sencilla sería calcular la media aritmética, pero no hemos llegado hasta aquí para hacer otra chapuza. La media aritmética otorga el mismo énfasis a todos los estudios, con lo que si tenemos algún estudio muy impreciso con resultados extremos nos distorsionará enormemente los resultados. Recordad que la media sigue siempre las colas de la distribución, viéndose muy influenciada por los valores extremos (lo que no le ocurre a su prima, la mediana).

Por esto tenemos que ponderar los estimadores de los distintos estudios. Esto podemos hacerlo de dos formas, teniendo en cuenta el número de sujetos de cada estudio, o bien, realizando una ponderación en base a los inversos de las varianzas de cada uno (ya sabéis, los cuadrados de los errores estándar). Esta última forma es la más compleja, así que es la que más gusta y más se utiliza. Claro que, como la matemática necesaria es cosa fina, se suelen utilizar programas especiales, ya sean módulos que funcionan dentro de los programas estadísticos como Stata, SPSS, SAS o R, o utilizando programas específicos como el famoso RevMan de la Colaboración Cochrane.

Como podéis ver, no me he quedado corto al tildar a la revisión

sistemática con metanálisis como la fiera corrupta de los diseños epidemiológicos. No obstante, tiene sus detractores. Todos conocemos a alguien que afirma no gustarle las revisiones sistemáticas porque casi todas acaban de la misma forma: “hacen falta más estudios de calidad para poder realizar recomendaciones con un grado de evidencia razonable”. Claro que, en estos casos, la culpa no es de las revisiones, sino de que no nos esmeramos lo suficiente con nuestros trabajos y la gran mayoría merecerían acabar en la máquina destructora de papel.

Otra polémica es la que tienen los que debaten sobre si es mejor una buena revisión sistemática o un buen ensayo clínico (pueden hacerse revisiones sobre otros tipos de diseños, incluyendo estudios observacionales). A mí esto me recuerda a la polémica sobre si deben hacer calimochos con buen vino o si es un pecado mezclar un buen vino con Coca-Cola. Polémicas aparte, si hay que tomar calimochos, os aseguro que con un buen vino estará mucho más rico, y algo parecido les pasa a las revisiones con la calidad de sus estudios primarios.

El problema de las revisiones sistemáticas es que, para que sean realmente útiles, hay que ser muy riguroso en su elaboración. Para que no se nos olvide nada, existen listas de recomendaciones y de verificación que nos permitan ordenar todo el procedimiento de creación y difusión de trabajos científicos sin que cometamos errores metodológicos u omisiones en el procedimiento.

Todo comenzó con un programa del Servicio de Salud del Reino Unido que terminó con la fundación de una iniciativa internacional para promover la transparencia y precisión de los trabajos de investigación biomédicos: la [red EQUATOR](#) (Enhancing the QUALity and Transparency Of health Research). Esta red se compone de expertos en metodología, comunicación y publicación, por lo que incluye profesionales implicados en la calidad de todo el proceso de producción y difusión de los resultados de investigación. Entre otros muchos objetivos, que podéis consultar en su página web, está el de diseñar un conjunto de recomendaciones para la realización y publicación de los diferentes tipos de estudios, lo que da lugar a las diferentes listas de verificación o declaraciones.

La lista de verificación diseñada para aplicar a las revisiones sistemáticas es la [declaración PRISMA](#) (Preferred Reporting Items for Systematic reviews and Meta-Analyses), que viene a sustituir a la declaración QUOROM (QUality Of Reporting Of Meta-analyses). Basándose en la definición de revisión sistemática de la Colaboración Cochrane, PRISMA nos ayuda a seleccionar, identificar y valorar los estudios incluidos en una revisión. Consta también de una lista de verificación y de un diagrama de flujo en el que se describe el paso de todos los trabajos considerados durante la realización de la revisión. Existe también una declaración menos

conocida para la valoración de metanálisis de estudios observacionales, la declaración MOOSE (Meta-analyses Of Observational Studies in Epidemiology).

La Colaboración Cochrane tiene también una metodología muy bien estructurada y definida, que podéis consultar en su página web. Esta es la razón por la que tienen tanto prestigio dentro del mundo de las revisiones sistemáticas, por estar hechas por profesionales que se dedican a ello siguiendo una metodología contrastada y rigurosa. De todas formas, incluso las revisiones Cochrane deben leerse de forma crítica y no darles nada por asegurado.

Y con esto hemos llegado al final por hoy. Quiero insistir en que el metanálisis debe hacerse siempre que sea posible y nos interese, pero asegurándonos previamente de que es correcto combinar los resultados. Si los estudios son muy heterogéneos no deberemos combinar nada, ya que los resultados que podríamos obtener tendrían una validez muy comprometida. Hay toda una serie de métodos y estadísticos para medir la homogeneidad o heterogeneidad de los estudios primarios, que influyen también en la forma en que analizaremos los datos combinados. Pero esa es otra historia...
