



Campos electromagnéticos y salud pública

Reproducción y radiofrecuencia: la evidencia científica más robusta revisada por la OMS

Un análisis sistemático de estudios en humanos y animales concluye que la evidencia experimental sobre exposición prenatal a **CEM-RF** no permite establecer una toxicidad reproductiva general.

Madrid, 23 de Septiembre de 2026

3. Reproducción

Este bloque temático aborda de forma integrada la evidencia sobre los posibles efectos de la exposición a CEM-RF en la reproducción, considerando tanto la fertilidad masculina como el embarazo, el desarrollo prenatal y los resultados al nacimiento. La salud reproductiva fue uno de los seis temas priorizados en el proceso promovido por la OMS para la evaluación de los posibles efectos de los CEM-RF sobre la salud (Verbeek et al., 2021). Dentro de este ámbito, la evidencia disponible plantea preguntas diferentes según el desenlace considerado y el tipo de estudio, por lo que la evaluación se articuló mediante revisiones complementarias de estudios observacionales en humanos y de estudios experimentales. En conjunto, este bloque comprende cuatro revisiones sistemáticas: dos sobre fertilidad masculina, una de estudios observacionales humanos y otra de estudios experimentales en mamíferos no humanos y espermatozoides humanos *in vitro*, y dos sobre embarazo y resultados reproductivos y del desarrollo, una observacional en humanos y otra experimental en mamíferos no humanos (Cordelli et al., 2024, 2023; Johnson et al., 2024; Kenny et al., 2024).

La separación en cuatro revisiones permite responder a preguntas distintas y, al mismo tiempo, complementarias. Los estudios observacionales humanos proporcionan evidencia directamente aplicable a la población general y trabajadora, pero están condicionados por las dificultades para caracterizar con precisión la exposición a CEM-RF, controlar la confusión y establecer relaciones dosis–respuesta a partir de indicadores que con frecuencia son aproximaciones de la dosis realmente absorbida. Los estudios experimentales permiten, por el contrario, controlar la exposición, utilizar grupos *sham* y explorar directamente diversos desenlaces reproductivos y mecanismos relacionados, aunque plantean problemas propios de extrapolación a humanos, especialmente cuando las exposiciones experimentales son superiores a las habituales en la población. En fertilidad masculina, este enfoque experimental se amplía además con estudios de espermatozoides humanos expuestos *in vitro*, que eliminan la extrapolación entre especies para determinados parámetros celulares, pero no reproducen la complejidad fisiológica de la exposición de un organismo completo. Esta complementariedad entre evidencia humana y experimental constituye, por tanto, un elemento central para interpretar el conjunto del bloque.

La primera revisión observacional evalúa la fertilidad masculina en adultos expuestos a CEM-RF procedentes de fuentes ambientales y ocupacionales, considerando exposiciones localizadas y de cuerpo entero (Kenny et al., 2024). Los desenlaces comprenden infertilidad masculina, concentración y recuento total de espermatozoides, morfología, motilidad total y progresiva y tiempo hasta el embarazo. En el protocolo, la SAR se consideró la métrica dosimétrica ideal para caracterizar la exposición; sin embargo, dado que las estimaciones individualizadas de SAR en los tejidos de interés rara vez están disponibles en los estudios epidemiológicos, se admitieron también medidas o modelos de campos electromagnéticos y densidad de potencia, así como diversos indicadores indirectos de exposición, entre ellos la duración y las características del uso del teléfono móvil. Además de sintetizar las asociaciones entre exposición y desenlace, la revisión planteó expresamente como objetivo secundario determinar si existían relaciones dosis–respuesta. Cuando la comparabilidad de los estudios lo permitió se realizaron metaanálisis dosis–respuesta; cuando no fue posible una síntesis cuantitativa, los resultados se integraron narrativamente.

La segunda revisión observacional examina los efectos de los CEM-RF sobre resultados reproductivos femeninos y del embarazo en adultos durante el periodo preconcepcional o la gestación (Johnson et al., 2024). Los desenlaces de interés incluyen parto prematuro, pequeño para la edad gestacional, incluido el bajo peso al nacer a término como indicador de restricción del crecimiento intrauterino, aborto espontáneo, muerte fetal y anomalías congénitas. Al igual que en la revisión de fertilidad masculina, se contemplan exposiciones ambientales y ocupacionales, localizadas o de cuerpo entero, y se considera la SAR como métrica ideal, aunque se admiten medidas y modelos de exposición y aproximaciones indirectas cuando aquella no está disponible. La síntesis combina, cuando los datos lo permiten, metaanálisis de efectos aleatorios comparando las categorías de mayor y menor exposición y análisis dosis-respuesta; cuando la heterogeneidad o la insuficiencia de datos impiden una combinación cuantitativa razonable, se recurre a síntesis narrativa.

La revisión experimental sobre fertilidad masculina integra dos modelos de evidencia: estudios in vivo en mamíferos no humanos y estudios de espermatozoides humanos expuestos in vitro (Cordelli et al., 2024). En los primeros se analizan desenlaces que abarcan desde la fertilidad propiamente dicha y la calidad seminal hasta toxicidad de los órganos reproductores y efectos hormonales; en los segundos se estudian directamente parámetros del espermatozoides humano tras una exposición controlada. La síntesis cuantitativa utiliza diferentes medidas de efecto en función del tipo de desenlace y explora la heterogeneidad mediante análisis por subgrupos, entre ellos especie, SAR y aumento de temperatura en los estudios animales, y estado de fertilidad de los donantes y SAR en los estudios in vitro. Esta revisión resulta especialmente relevante para distinguir entre una asociación epidemiológica y un efecto observado bajo condiciones experimentales controladas, pero también exige considerar cuidadosamente la dosimetría y la aplicabilidad a exposiciones humanas reales, dado que una parte importante de los experimentos utilizó niveles superiores a los habituales en la población.

La cuarta revisión analiza experimentalmente la exposición a CEM-RF durante la gestación en mamíferos no humanos y sus posibles efectos sobre la fecundidad, el embarazo, la salud de la descendencia al nacimiento y determinados efectos posteriores del desarrollo (Cordelli et al., 2023). Su PECO define como población a mamíferos expuestos exclusivamente *in utero*; como exposición, CEM-RF entre 100 kHz y 300 GHz, además de pulsos electromagnéticos considerados separadamente. Como comparadores, animales sometidos a exposición simulada (*sham*) o, cuando procede, controles sometidos al mismo incremento térmico mediante calentamiento directo y como desenlaces, pérdidas preimplantacionales, fetos muertos o reabsorbidos, tamaño de camada, peso y longitud fetal, malformaciones, proporción de sexos y determinados efectos tardíos en la descendencia. La inclusión de controles térmicos tiene particular interés porque permite abordar una cuestión esencial en este campo: diferenciar los efectos atribuibles al calentamiento, ya conocidos para exposiciones suficientemente intensas, de posibles efectos asociados a los CEM-RF en otras condiciones de exposición.

Las cuatro revisiones se desarrollaron a partir de protocolos previamente establecidos y dentro del marco metodológico impulsado por la OMS para estas evaluaciones. Las dos revisiones observacionales humanas de fertilidad masculina y resultados adversos del embarazo comparten un protocolo publicado (Kenny et al., 2022) y registros específicos en PROSPERO (SR3A, CRD42021265401; SR3B, CRD42021266268), y aplican el WHO Handbook y las recomendaciones COSTER. Incluyen búsquedas amplias en bases bibliográficas y EMF-Portal, literatura gris y búsqueda complementaria de referencias, selección y extracción con participación de revisores independientes, evaluación del riesgo de sesgo mediante OHAT y valoración de la certeza mediante OHAT-GRADE.

En ambos casos, la caracterización de la exposición constituye un aspecto metodológico particularmente crítico, debido a la frecuente utilización de variables indirectas, como tiempo de uso del teléfono o distancia a una fuente, en lugar de estimaciones individualizadas de la energía de RF realmente absorbida. Las dos revisiones experimentales proceden asimismo de un protocolo común publicado previamente (Pacchierotti et al., 2021) y siguen las recomendaciones de la OMS y del National Toxicology Program/Office of Health Assessment and Translation (NTP/OHAT) para revisiones sistemáticas de evidencia experimental. La revisión de embarazo y desarrollo fue registrada en PROSPERO, mientras que la revisión experimental de fertilidad masculina utilizó PROSPERO para los estudios animales y Open Science Framework para los experimentos con espermatozoides humanos in vitro.

Ambas incorporan evaluación formal del riesgo de sesgo, metaanálisis cuando la homogeneidad de los datos lo permite, análisis de heterogeneidad y de variables relacionadas con la exposición y valoración de la certeza mediante GRADE. En este componente experimental adquieren especial relevancia la caracterización y dosimetría de la exposición, el cegamiento de la evaluación de los desenlaces, la distinción entre estudios con mayor o menor preocupación por riesgo de sesgo y, para interpretar la relevancia para humanos, la magnitud de las exposiciones empleadas y su posible componente térmico.

Manteniendo la trazabilidad PECO y la separación entre evidencia epidemiológica y experimental, hemos organizado este bloque en cuatro subapartados consecutivos basados en las cuatro revisiones, como se ha indicado: fertilidad masculina en estudios observacionales humanos, resultados reproductivos femeninos y del embarazo en estudios observacionales humanos, fertilidad masculina en estudios experimentales con mamíferos no humanos y espermatozoides humanos in vitro y, por último, embarazo y resultados del desarrollo en estudios experimentales con mamíferos no humanos.

Como se ha indicado, la lectura cruzada de las cuatro revisiones permite contrastar la evidencia directamente obtenida en humanos con la procedente de sistemas experimentales y distinguir, al mismo tiempo, entre diferentes niveles del proceso reproductivo: parámetros seminales y capacidad reproductiva masculina, evolución del embarazo y resultados perinatales, y alteraciones observadas bajo exposición prenatal controlada.

Esta integración es especialmente importante porque las principales fuentes de incertidumbre no son idénticas en los distintos diseños: en epidemiología predominan los problemas de caracterización de la exposición, confusión, heterogeneidad de las fuentes y uso de indicadores indirectos de dosis; en los estudios experimentales resultan centrales el riesgo de sesgo, la calidad de la dosimetría, la heterogeneidad entre protocolos y la extrapolación desde niveles de exposición experimentales, en ocasiones muy superiores a los relevantes para la población, hacia condiciones humanas reales.

En términos de evaluación del riesgo, el valor del bloque reside precisamente en examinar conjuntamente estas líneas de evidencia sin equipararlas de forma automática: una asociación observacional no demuestra por sí misma causalidad, del mismo modo que un efecto experimental a dosis elevadas no implica necesariamente un riesgo a los niveles de exposición habituales. Las cuatro revisiones proporcionan así un marco metodológicamente estructurado para valorar, con certeza graduada y atendiendo a la relación dosis-respuesta y a la relevancia de las condiciones de exposición, qué puede afirmarse y qué permanece incierto sobre los posibles efectos de los CEM-RF en la reproducción.

3.1. Fertilidad masculina: estudios observacionales en humanos

3.1.1. Referencia completa del artículo

Título: *The effects of radiofrequency exposure on male fertility: A systematic review of human observational studies with dose-response meta-analysis.* Kenny R.P.W., Johnson E.E., Adesanya A.M., Richmond C., Beyer F., Calderon C., Rankin J., Pearce M.S., Toledano M., Craig D., Pearson F.. **Revista y datos bibliográficos:** *Environment International*, 190, 108817 (2024). DOI: 10.1016/j.envint.2024.108817. **Registro de protocolo:** PROSPERO CRD42021265401.

3.1.2. Objetivos y preguntas de investigación

El objetivo principal de esta revisión sistemática fue evaluar y sintetizar la evidencia procedente de estudios observacionales en humanos sobre los posibles efectos de la exposición a CEM-RF en la fertilidad masculina. Los autores señalan que, aunque trabajos previos habían examinado posibles efectos sobre parámetros reproductivos masculinos, la evidencia disponible seguía siendo poco clara y las conclusiones anteriores procedían de revisiones no sistemáticas. Además, no existía hasta ese momento ninguna revisión sistemática de estudios observacionales que evaluase conjuntamente los efectos sobre los resultados reproductivos masculinos de la exposición localizada, como la asociada a teléfonos móviles, y de la exposición de cuerpo entero, como la procedente de fuentes ambientales u ocupacionales.

La pregunta de investigación se formuló explícitamente mediante el marco Población, Exposición, Comparador, Resultado (PECO): en estudios observacionales en humanos, ¿cuáles son los efectos de la exposición localizada o de cuerpo entero a CEM-RF (E), en comparación con una exposición nula o de bajo nivel (C), en varones adultos (P), sobre la infertilidad masculina, la morfología, la motilidad, la concentración o el recuento espermáticos y el tiempo hasta el embarazo (O)? De este modo, la revisión abarcó tanto diferentes parámetros de calidad seminal como desenlaces reproductivos de mayor alcance, concretamente la infertilidad masculina y el tiempo hasta el embarazo.

Como objetivo secundario, los autores se propusieron determinar si existía una relación dosis-respuesta entre la exposición a CEM-RF y los resultados de fertilidad masculina considerados. Por tanto, la revisión no se limitó a examinar la existencia de asociaciones entre exposición y desenlace, sino que evaluó, cuando los datos disponibles lo permitieron, si las variaciones en el nivel de exposición se relacionaban con cambios cuantificables en los parámetros reproductivos.

3.1.3. Metodología

Se desarrolló y siguió un protocolo detallado (Kenny et al., 2022) y registrado en PROSPERO (CRD42021265401), correspondiente a SR3A. Este protocolo fue desarrollado conjuntamente para las revisiones observacionales sobre fertilidad masculina (SR3A) y resultados reproductivos femeninos (SR3B), aunque el presente artículo aborda exclusivamente la primera. La planificación metodológica siguió el *WHO Handbook for Guideline Development* y las recomendaciones COSTER, y el protocolo se elaboró conforme a PRISMA-P.

Búsqueda y fuentes

La búsqueda bibliográfica se diseñó con un enfoque amplio y sensible. Las principales fuentes fueron MEDLINE, Embase y EMF-Portal. La estrategia se desarrolló iterativamente a partir de los conceptos centrales de la pregunta de investigación, combinando términos libres y encabezamientos temáticos e incorporando desenlaces identificados con la participación de expertos clínicos. No se aplicaron restricciones por fecha ni por idioma. La búsqueda inicial se realizó en noviembre de 2020 y se actualizó en diciembre de 2021 y enero de 2023. Los resultados se gestionaron mediante EndNote, donde se eliminaron los duplicados antes del cribado.

La búsqueda se complementó con literatura gris. En abril de 2021 se consultó OpenGrey y se realizaron búsquedas de guías e informes de organismos de salud pública y protección radiológica, tesis y comunicaciones científicas. También se utilizaron Web of Science, para resúmenes de congresos, e IEEE Xplore®. Estas búsquedas se actualizaron en enero de 2022 y enero de 2023 y, tras dejar de estar disponible OpenGrey, se recurrió a búsquedas avanzadas en Google. Además, se examinaron las referencias de revisiones sistemáticas y narrativas relevantes, las referencias y citas posteriores de los estudios incluidos y trabajos adicionales identificados por expertos en CEM-RF. La última actualización aportó un único estudio adicional.

Criterios de elegibilidad

Los criterios de elegibilidad se establecieron a priori mediante el marco PECO. La población de interés estuvo constituida por varones adultos expuestos a CEM-RF procedentes de fuentes ambientales, ocupacionales o no ocupacionales, en los que se evaluarán desenlaces relacionados con la fertilidad masculina.

Se consideraron tanto exposiciones localizadas como de cuerpo entero. La tasa de absorción específica (SAR), expresada en W/kg, se estableció como la métrica ideal, aunque, dada la dificultad de disponer de estimaciones de SAR en los órganos reproductores en estudios epidemiológicos, se admitieron también medidas indirectas o *proxies* de exposición basadas en niveles medidos o modelizados de campo eléctrico o magnético, densidad de potencia y otras aproximaciones comparables. Para telefonía móvil se aceptaron medidas autoinformadas, como duración o frecuencia de uso, y medidas objetivas. Para estaciones base se exigieron estimaciones objetivas de la exposición o de la distancia a la fuente, excluyéndose la distancia exclusivamente autoinformada. También se contemplaron mediciones puntuales, exposímetros personales y modelos de predicción.

En el ámbito ocupacional se aceptaron estimaciones basadas en mediciones, observación, valoración experta, combinaciones de estos procedimientos o matrices empleo-exposición (*Job-Exposure Matrix*, JEM). Se excluyeron las clasificaciones basadas exclusivamente en el título o categoría profesional, así como medidas excesivamente imprecisas de exposición, por ejemplo, la mera posesión alguna vez de un teléfono móvil. También se excluyeron las exposiciones derivadas de tecnologías médicas cuando los sujetos estudiados eran pacientes, en lugar de trabajadores que utilizaban habitualmente esos equipos.

Los comparadores podían corresponder a ausencia o bajo nivel de exposición frente a niveles superiores, a dos o más categorías de exposición o duración, o a datos continuos que permitieran evaluar una relación dosis-respuesta. Los desenlaces de interés comprendieron infertilidad masculina, tiempo hasta el embarazo, concentración espermática, recuento espermático total, morfología y motilidad. Los casos de infertilidad masculina debían basarse en diagnóstico médico, y para los parámetros seminales se tomaron como referencia los rangos establecidos por la OMS. Se excluyeron otros parámetros espermáticos cuya validez diagnóstica para infertilidad no estuviera establecida y los estudios basados en infertilidad masculina exclusivamente autoinformada.

Fueron elegibles estudios de cohortes y de casos y controles. Se excluyeron los estudios transversales sobre infertilidad masculina, al no permitir establecer adecuadamente la temporalidad entre exposición y desenlace, los estudios con autoselección de participantes a partir de una población de origen no definida, los informes de casos y los estudios preclínicos o *in vitro*. No se establecieron restricciones por año o idioma y se contempló tanto literatura publicada como no publicada.

Selección y extracción de datos

La selección se realizó en dos fases, primero mediante cribado de títulos y resúmenes y posteriormente mediante evaluación de los textos completos potencialmente elegibles. Pares de revisores realizaron las evaluaciones de forma independiente, y las discrepancias se resolvieron mediante discusión o, cuando fue necesario, mediante un tercer revisor. El mismo procedimiento se aplicó a la literatura gris.

La extracción de datos se realizó mediante formularios estandarizados y previamente pilotados. Un revisor efectuó la extracción inicial y un segundo comprobó su integridad y exactitud frente a las publicaciones originales. Las discrepancias se resolvieron por consenso o mediante un tercer revisor. Cuando faltaba información relevante, se contactó con los autores de los estudios primarios, aunque estos contactos no proporcionaron finalmente datos adicionales.

Medidas de efecto y síntesis cuantitativa

La estrategia analítica prevista contemplaba metaanálisis pareados y análisis dosis-respuesta. Inicialmente se consideró utilizar diferencias de medias estandarizadas (SMD) para combinar estudios que presentaban los mismos desenlaces mediante métricas diferentes. Sin embargo, los autores descartaron esta opción al no poder asumir que las diferencias entre estudios obedecieran exclusivamente a las escalas utilizadas y no a diferencias en la fiabilidad de las medidas o en la variabilidad de las poblaciones.

También se valoró combinar mediante diferencias de medias (MD) estudios que comparaban distintos niveles de exposición. La heterogeneidad de las categorías empleadas hizo que los autores considerasen poco plausible una síntesis convencional de exposición alta frente a baja o nula. Por ello, cuando existieron datos suficientes, se priorizaron los metaanálisis dosis-respuesta, considerados más interpretables y con menor heterogeneidad.

Los análisis dosis-respuesta se realizaron con el paquete *dosresmeta* en R cuando al menos dos estudios evaluaban un mismo desenlace mediante una cuantificación comparable de la exposición. Se probaron modelos lineales y cuadráticos no lineales, evaluándose su ajuste mediante χ^2 , *log-likelihood*, criterio de información de Akaike (AIC) y criterio de información bayesiano (BIC). En todos los análisis el mejor ajuste correspondió al modelo lineal, por lo que se utilizaron modelos lineales de efectos aleatorios. La exposición se expresó fundamentalmente mediante el tiempo de uso o conversación con teléfono móvil y los efectos se estimaron como MD por incremento de una hora diaria de uso.

Cuando los estudios presentaban categorías de exposición, se asignaron valores continuos representativos para poder incorporarlas al análisis dosis-respuesta. Los coeficientes de regresión β pudieron combinarse con las MD cuando eran coeficientes no estandarizados. La heterogeneidad entre estudios se cuantificó mediante τ , estimado por máxima verosimilitud restringida (REML), e I^2 .

El escaso número de estudios impidió realizar análisis de subgrupos y análisis de sensibilidad. Cuando los datos no permitieron metaanálisis, se efectuó una síntesis narrativa siguiendo las recomendaciones SWiM (*Synthesis Without Meta-analysis*). Tampoco pudieron realizarse las pruebas estadísticas previstas para evaluar el sesgo de publicación debido al reducido número de estudios disponibles.

Evaluación del riesgo de sesgo (RoB)

El riesgo de sesgo se evaluó mediante la herramienta OHAT del National Toxicology Program, aplicada a nivel de estudio y desenlace. Se consideraron siete dominios: selección o participación, medición de la exposición, evaluación del desenlace, confusión, pérdidas o exclusiones, notificación selectiva y otras amenazas potenciales a la validez interna. Cada dominio se clasificó como riesgo definitivamente bajo, probablemente bajo, probablemente alto o definitivamente alto.

La evaluación fue realizada independientemente por dos revisores, con resolución de discrepancias mediante consenso y, cuando fue necesario, intervención de un tercer revisor. Para el dominio de confusión se identificaron como factores críticos la edad, la etnia, el índice de masa corporal, el nivel socioeconómico, el tabaquismo y el consumo de alcohol. También se consideraron otros factores potencialmente relevantes, como localización geográfica, coexposiciones ocupacionales, calor, ruido ambiental y contaminación atmosférica. La ausencia de ajuste por confusión no determinaba la exclusión del estudio, pero sí influía en la valoración de su riesgo de sesgo.

Certeza de la evidencia (OHAT-GRADE)

La certeza de la evidencia se evaluó mediante OHAT-GRADE para los desenlaces en los que fue posible realizar un metaanálisis dosis-respuesta. Se consideraron los dominios de riesgo de sesgo, inconsistencia, indirectitud, imprecisión y sesgo de publicación, junto con los factores que podían aumentar la certeza, como magnitud del efecto, confusión plausible y existencia de un gradiente dosis-respuesta. No se aplicó el dominio adicional de consistencia entre diferentes modelos o diseños, decisión prevista en el protocolo tras consultar su pertinencia con expertos en OHAT y GRADE.

El nivel inicial de certeza se estableció para cada desenlace y posteriormente se aplicaron las correspondientes degradaciones o mejoras en función de los dominios evaluados. Un revisor realizó la valoración inicial, comprobada por un segundo, y posteriormente dos revisores evaluaron independientemente cada dominio, resolviendo las discrepancias por consenso o mediante un tercer revisor. La calificación final se expresó como alta, moderada, baja o muy baja, y las conclusiones se formularon utilizando la terminología recomendada por Santesso et al. Para los desenlaces que solo pudieron sintetizarse narrativamente no se realizó una evaluación OHAT-GRADE completa, considerándose en su lugar el riesgo de sesgo del conjunto de estudios disponibles.

3.1.4. Resultados

Flujo PRISMA

El proceso de búsqueda y selección se realizó conjuntamente para las dos revisiones observacionales sobre reproducción, SR3A, centrada en fertilidad masculina, y SR3B, sobre resultados reproductivos femeninos. El diagrama PRISMA muestra 22.352 registros cribados tras eliminar 7.545 duplicados. De ellos, 284 informes fueron seleccionados para recuperación en texto completo, nueve no pudieron obtenerse y 275 fueron finalmente evaluados para determinar su elegibilidad. Paralelamente, el encadenamiento de citas permitió identificar 41 registros adicionales y la consulta a expertos aportó otros dos, de modo que se evaluaron 43 informes adicionales procedentes de estas vías.

En el conjunto de ambas revisiones se incluyeron 29 estudios. Nueve correspondieron específicamente a fertilidad masculina, siete realizados en población general y dos en poblaciones ocupacionales. El artículo contiene una inconsistencia interna en este punto, ya que la sección 4.1 menciona inicialmente ocho estudios, seis de población general y dos ocupacionales, pero el resumen, la descripción posterior de los estudios, las tablas, la síntesis y la discusión trabajan de forma consistente con nueve estudios, siete de población general y dos ocupacionales. Este último es, por tanto, el conjunto de evidencia efectivamente analizado en la revisión.

Características de los estudios incluidos

Los siete estudios de población general fueron Chen et al. (2022), Feijo et al. (2011), Hatch et al. (2021), Jurewicz et al. (2014), Lewis et al. (2017), Rago et al. (2013) y Zhang et al. (2016). Los dos estudios ocupacionales fueron Lancranjan et al. (1975), en técnicos expuestos a microondas, y Ye et al. (2007), en marinos expuestos a radar. Los estudios se realizaron en China, Estados Unidos, Dinamarca, Brasil, Polonia, Rumanía e Italia; Hatch et al. incluyó participantes tanto de Estados Unidos como de Dinamarca. El tamaño muestral fue muy variable, desde 31 participantes en Lancranjan et al. hasta 3.100 participantes en Hatch et al.

Ocho de los nueve estudios fueron tratados como estudios de cohortes, mientras que Lancranjan et al. (1975) parecía corresponder a un estudio de casos y controles, aunque los propios autores señalan que la descripción de su diseño era poco clara. Los participantes fueron predominantemente adultos jóvenes y de mediana edad, aunque la información sobre edad, nivel educativo, consumo de alcohol, tabaquismo y periodo de abstinencia sexual fue heterogénea o incompleta entre estudios.

En los estudios de población general, la fuente de exposición predominante fue el teléfono móvil. Seis estudios evaluaron exclusivamente exposición asociada al teléfono y uno analizó además ordenadores y otros dispositivos eléctricos. Las métricas empleadas fueron muy diversas: minutos u horas de utilización diaria, tiempo de conversación, años de uso, tráfico mensual de datos, tiempo durante el cual el teléfono se mantenía en el bolsillo delantero del pantalón o simplemente la localización habitual del dispositivo. La información sobre la distribución de la exposición, los niveles de exposición de fondo, la intensidad, la localización corporal y las posibles coexposiciones fue escasa, incompleta o heterogénea entre los estudios.

La SAR, establecida a priori como la métrica ideal, no fue utilizada por ninguno de los estudios incluidos. La exposición se caracterizó fundamentalmente mediante *proxies* de uso, en particular duración del uso del teléfono o tiempo de conversación, lo que constituye una limitación relevante dado que estas métricas no cuantifican directamente la exposición de los órganos reproductores.

Resultados cuantitativos en población general

Los análisis cuantitativos se concentraron en la duración del uso del teléfono móvil. Cuando al menos dos estudios proporcionaban una cuantificación comparable de la exposición y del desenlace, los autores realizaron metaanálisis dosis-respuesta mediante modelos lineales de efectos aleatorios. Los resultados se expresaron como diferencia media asociada a una hora adicional de utilización diaria.

Para la concentración espermática, cinco estudios evaluaron la exposición a teléfonos móviles y uno adicional el uso de teléfonos, ordenadores y otros dispositivos eléctricos. Tres estudios pudieron incluirse en el metaanálisis dosis-respuesta, con niveles de utilización comprendidos entre 0 y 300 minutos diarios y un total de 713 participantes. La diferencia media estimada por cada hora adicional de uso diario fue de $1,6 \times 10^6$ espermatozoides/mL (IC95 % -1,7 a 4,9; $p = 0,34$). La estimación fue compatible con ausencia de efecto y la heterogeneidad fue elevada ($I^2 = 89,9$ %; $\tau = 0,0472$). La certeza de la evidencia se calificó como muy baja.

Dos estudios adicionales evaluaron la concentración mediante métricas distintas y no pudieron incorporarse al metaanálisis. Jurewicz et al. no encontró diferencias estadísticamente significativas entre 11-25 años y 0-5 años de utilización del teléfono ($\beta = -0,33$; $p = 0,14$). Chen et al. tampoco observó diferencias significativas al comparar 3,1-4,5 horas o más de 4,5 horas diarias con tres horas o menos, con cambios porcentuales de $-1,2\%$ (IC95 % $-10,2$ a $8,7$) y $-2,5\%$ (IC95 % $-10,5$ a $6,2$), respectivamente.

Para la morfología espermática, tres estudios con 713 participantes pudieron combinarse en el análisis dosis-respuesta. La diferencia media por cada hora adicional de utilización diaria fue de 0,15 puntos porcentuales (IC95 % $-0,21$ a $0,51$; $p = 0,41$), con heterogeneidad elevada ($I^2 = 83,3\%$; $\tau = 0,0044$). La evidencia fue calificada igualmente como de certeza muy baja.

Jurewicz et al. no pudo integrarse en este análisis porque informó anomalías de cabeza, cuello y cola en lugar de porcentaje de formas normales. En este estudio, utilizar el teléfono durante más de 11 años, frente a 0-5 años, se asoció con un mayor porcentaje de espermatozoides atípicos ($\beta = 19$; $p = 0,002$) y de anomalías de la cabeza ($\beta = 17,58$; $p = 0,01$), pero no con anomalías del cuello o de la cola. Tampoco se observaron esas asociaciones al comparar 6-10 años con 0-5 años de uso.

Para la motilidad total, cuatro estudios proporcionaron datos, pero ninguno fue suficientemente comparable como para realizar una síntesis cuantitativa; entre los resultados comunicados individualmente, Lewis et al. no observó diferencias estadísticamente significativas entre no usuarios y hombres que utilizaban el teléfono menos de dos horas diarias ($\beta = 6,38$ puntos porcentuales; IC95 % $-1,2$ a 14) o dos horas o más ($\beta = 3,59$; IC95 % $-6,61$ a $13,79$). Chen et al. tampoco detectó diferencias significativas al comparar 3,1-4,5 horas o más de 4,5 horas frente a tres horas o menos. En cambio, Jurewicz et al. encontró una reducción estadísticamente significativa de la motilidad al comparar 11-25 años con 0-5 años de utilización ($\beta = -11,08$ puntos porcentuales; $p = 0,02$). En conjunto, por tanto, los resultados fueron inconsistentes y no permitieron establecer un patrón claro.

Para la motilidad progresiva, dos estudios con 560 participantes pudieron combinarse. La diferencia media por cada hora adicional de uso diario fue de $-0,46$ puntos porcentuales (IC95 % $-1,04$ a $0,13$; $p = 0,1307$). A diferencia de otros desenlaces, no se observó heterogeneidad estadística ($I^2 = 0\%$; $\tau = 0$). Aunque la estimación puntual se orientaba hacia una ligera reducción, el intervalo de confianza incluía la ausencia de efecto y la certeza de la evidencia fue muy baja.

Chen et al., no incluido en ese metaanálisis, tampoco observó diferencias significativas en motilidad progresiva para 3,1-4,5 horas de uso diario frente a tres horas o menos, cambio de $1,2\%$ (IC95 % $-3,8$ a $6,4$), ni para más de 4,5 horas, cambio de $-0,9\%$ (IC95 % $-5,3$ a $3,7$).

Para el recuento espermático total, dos estudios con 216 participantes pudieron incorporarse al metaanálisis dosis-respuesta. La diferencia media estimada por cada hora adicional de uso diario fue de $-0,44 \times 10^6$ espermatozoides por eyaculado (IC95 % $-2,59$ a $1,70$; $p = 0,6833$). La heterogeneidad fue muy elevada ($I^2 = 90\%$; $\tau = 0,0246$) y la certeza de la evidencia se calificó como muy baja.

Chen et al., no integrable en el metaanálisis, tampoco observó diferencias estadísticamente significativas en el recuento total al comparar 3,1-4,5 horas o más de 4,5 horas de uso diario con tres horas o menos, con cambios de $-7,6\%$ (IC95 % $-18,2$ a $4,3$) y $-5,6\%$ (IC95 % $-15,4$ a $5,2$), respectivamente.

En conjunto, los cuatro metaanálisis dosis-respuesta no mostraron asociaciones estadísticamente significativas entre una mayor duración diaria de uso del teléfono móvil y concentración, morfología, motilidad progresiva o recuento espermático total. Sin embargo, la ausencia de asociaciones significativas no se interpretó como evidencia sólida de ausencia de efecto, dado que la certeza fue muy baja en todos los casos.

Posición del teléfono y tiempo hasta el embarazo

Varios estudios evaluaron si llevar el teléfono móvil cerca de los órganos reproductores, particularmente en el bolsillo delantero del pantalón, se asociaba con cambios en los parámetros seminales. En términos generales, los autores consideran que llevar el teléfono en esa posición puede tener poco o ningún efecto sobre concentración espermática, recuento total, morfología, motilidad progresiva o tiempo hasta el embarazo.

Para concentración espermática, Hatch et al. no encontró diferencias estadísticamente significativas al comparar llevar el teléfono en el bolsillo durante 3-7 horas o durante ocho horas o más frente a 0-2 horas. Chen et al. tampoco observó un efecto significativo de llevar el teléfono en el pantalón, y Lewis et al. y Rago et al. no detectaron diferencias significativas según la localización del dispositivo.

Para morfología, Rago et al. y Lewis et al. tampoco observaron diferencias significativas entre llevar el teléfono en el pantalón y otras localizaciones o no utilizarlo. En motilidad total, Hatch et al., Chen et al. y Lewis et al. informaron asimismo resultados compatibles con poca o ninguna diferencia. Lewis et al. sí observó un aumento estadísticamente significativo del recuento total de espermatozoides móviles en quienes llevaban el teléfono en el bolsillo del pantalón frente a los no usuarios ($\beta = 1,65$; IC95 % 1,02 a 2,67), hallazgo aislado cuya interpretación se ve limitada por el riesgo de sesgo del estudio.

El tiempo hasta el embarazo fue evaluado únicamente por Hatch et al. Comparados con hombres que no llevaban el teléfono en el bolsillo delantero, los *fecundability ratios* ajustados fueron 0,89 (IC95 % 0,78 a 1,02) para menos de 1-2 horas, 1,00 (IC95 % 0,88 a 1,14) para 3-7 horas y 0,92 (IC95 % 0,80 a 1,06) para ocho horas o más. Ninguna de estas estimaciones mostró una asociación estadísticamente significativa.

Otras fuentes de CEM-RF en población general

La evidencia sobre fuentes distintas de la telefonía móvil fue muy escasa. Un único estudio, Chen et al., evaluó el uso de ordenadores y otros dispositivos eléctricos además de teléfonos móviles. Sus resultados sugirieron poca o ninguna diferencia en concentración, motilidad total o recuento espermático total. Sin embargo, este estudio fue considerado con riesgo probablemente alto tanto en caracterización de la exposición como en evaluación del desenlace.

Por tanto, la evidencia disponible en población general se concentra casi exclusivamente en el uso del teléfono móvil y apenas permite evaluar otras fuentes de exposición.

Resultados en estudios ocupacionales

Los dos estudios ocupacionales identificados no pudieron combinarse cuantitativamente debido a las diferencias en población, fuente de exposición y evaluación de los desenlaces. Lancranjan et al. estudió técnicos expuestos a microondas, mientras que Ye et al. evaluó marinos expuestos a equipos de radar.

Para la concentración espermática, ambos estudios informaron valores inferiores en algunos grupos expuestos. En Lancranjan et al., la concentración media fue de $50 \times 10^6/\text{mL}$ (DE 24) en los técnicos expuestos y de $60 \times 10^6/\text{mL}$ (DE 34) en los no expuestos, con $p < 0,02$. En Ye et al., los valores medios en los cinco grupos de marinos expuestos oscilaron entre $22,06 \times 10^6/\text{mL}$ (DE 2,38) y $25,16 \times 10^6/\text{mL}$ (DE 2,07), mientras que los dos grupos no expuestos presentaron valores de $32,61 \times 10^6/\text{mL}$ (DE 2,94) y $22,61 \times 10^6/\text{mL}$ (DE 2,34).

Para la morfología, únicamente Lancranjan et al. proporcionó resultados. Los técnicos expuestos presentaron una media del 70 % de formas normales (DE 6), frente al 82 % (DE 7) en los controles no expuestos, con $p < 0,001$.

Para la motilidad, ambos estudios comunicaron diferencias estadísticamente significativas. En Lancranjan et al., la motilidad media fue del 36 % (DE 10) en los técnicos expuestos y del 54 % (DE 19) en los no expuestos, con $p < 0,001$. En Ye et al., la motilidad media en los grupos expuestos osciló entre 31,22 % (DE 2,26) y 47,2 % (DE 1,95), mientras que en los dos grupos no expuestos fue de 76,8 % (DE 1,59) y 40,97 % (DE 1,96); las diferencias entre grupos fueron estadísticamente significativas, con $p < 0,05$.

Ningún estudio ocupacional informó sobre recuento espermático total ni sobre tiempo hasta el embarazo. Pese a las diferencias observadas en algunos parámetros seminales, los autores consideran que el efecto de la exposición ocupacional a CEM-RF sobre la fertilidad masculina es incierto, debido al escaso número de estudios, la heterogeneidad entre fuentes de exposición y poblaciones y el elevado riesgo de sesgo.

Resultados cualitativos y consistencia global

La síntesis narrativa muestra un cuerpo de evidencia heterogéneo. En población general, la mayoría de los estudios evaluaron distintas aproximaciones al uso del teléfono móvil, pero las diferencias en duración, posición, años de uso y definición de la exposición impidieron integrar muchos resultados. Algunas asociaciones individuales alcanzaron significación estadística, como la menor motilidad con más de 11 años de uso en Jurewicz et al. o determinadas anomalías morfológicas en ese mismo estudio, mientras que otros trabajos no reprodujeron esos hallazgos.

Asimismo, los estudios ocupacionales informaron diferencias en algunos parámetros seminales entre grupos expuestos y no expuestos, pero procedían únicamente de dos investigaciones muy diferentes entre sí y con limitaciones metodológicas importantes. La revisión no identifica, por tanto, un patrón consistente que permita relacionar de forma fiable una mayor exposición a CEM-RF con un deterioro sistemático de los parámetros de fertilidad masculina.

Evaluación del riesgo de sesgo

El riesgo de sesgo fue una limitación importante del conjunto de evidencia. Los nueve estudios incluidos presentaron riesgo probablemente o definitivamente alto en al menos dos dominios de la herramienta OHAT.

La caracterización de la exposición fue uno de los problemas principales. Ningún estudio fue clasificado como definitivamente bajo riesgo en este dominio; solo Feijo et al. y Hatch et al. fueron considerados probablemente bajos. Seis estudios fueron clasificados como probablemente altos y Jurewicz et al. como definitivamente alto. En total, siete de los nueve estudios presentaron riesgo probablemente o definitivamente alto de sesgo en la caracterización de la exposición.

También se observaron problemas importantes en la evaluación del desenlace. Solo Jurewicz et al. y Lewis et al. fueron clasificados como definitivamente bajos en este dominio, mientras que los restantes siete estudios fueron considerados probablemente altos.

En relación con la confusión, Lancranjan et al., Lewis et al. y Ye et al. presentaron riesgo probablemente alto. También se identificaron problemas de notificación selectiva y pérdidas en varios estudios. Tres trabajos, Feijo et al., Lancranjan et al. y Ye et al., fueron considerados inadecuados en cuanto a los métodos estadísticos utilizados, fundamentalmente por insuficiente control de factores de confusión u otros problemas analíticos.

Las limitaciones fueron especialmente acusadas en Lancranjan et al., clasificado con riesgo probablemente o definitivamente alto en todos los dominios evaluados. En general, los autores destacan que la caracterización de la exposición fue inconsistente y poco detallada y que ninguno de los estudios utilizó SAR como métrica de exposición. En su lugar, predominaban *proxies* como tiempo de uso del teléfono, cuya relación con la exposición real de los órganos reproductores es indirecta.

Certeza de la evidencia (OHAT-GRADE)

La certeza se evaluó formalmente mediante OHAT-GRADE para los cuatro desenlaces de población general que pudieron analizarse mediante metaanálisis dosis-respuesta: concentración espermática, morfología, motilidad progresiva y recuento espermático total. En los cuatro casos, la certeza final fue muy baja.

Los estudios de cohorte partieron de un nivel inicial de certeza moderada. Para la concentración espermática, la certeza se degradó por riesgo de sesgo, inconsistencia e indirectitud. La inconsistencia fue particularmente importante, dada la elevada heterogeneidad estadística ($I^2 = 89,9 \%$). La indirectitud se relacionó con el hecho de que el tiempo de conversación o de uso del teléfono constituye un *proxy* de exposición que puede no representar adecuadamente la dosis recibida por los órganos reproductores.

La morfología espermática fue degradada por razones similares, incluyendo riesgo de sesgo, elevada inconsistencia ($I^2 = 83,3 \%$) e indirectitud. Para la motilidad progresiva no se observó heterogeneidad estadística, por lo que no se degradó por inconsistencia, pero sí por riesgo de sesgo e indirectitud. El recuento espermático total también fue degradado por riesgo de sesgo, inconsistencia, con $I^2 = 90 \%$, e indirectitud.

No se aplicaron mejoras de certeza en ninguno de los cuatro desenlaces. En conjunto, la revisión considera que la evidencia disponible es muy incierta acerca del efecto del uso del teléfono móvil sobre concentración, morfología, motilidad progresiva y recuento espermático total.

Para los resultados que no pudieron someterse a análisis dosis-respuesta, incluidos motilidad total, tiempo hasta el embarazo y los estudios ocupacionales, no se realizó una evaluación OHAT-GRADE formal completa. Su interpretación se basó fundamentalmente en el riesgo de sesgo de los estudios y en la consistencia de los resultados disponibles.

En síntesis, los resultados cuantitativos y narrativos no proporcionan evidencia suficientemente sólida para establecer si la exposición a CEM-RF afecta o no a la fertilidad masculina. Los metaanálisis dosis-respuesta no mostraron asociaciones estadísticamente significativas para los cuatro parámetros seminales evaluados, pero la certeza de la evidencia fue muy baja. Las asociaciones observadas en algunos estudios individuales, tanto en población general como ocupacional, se produjeron en un contexto de elevada heterogeneidad metodológica, caracterización indirecta de la exposición y riesgo de sesgo. Por ello, los autores consideran que no puede confiarse en que el cuerpo actual de evidencia permita determinar el efecto de los CEM-RF sobre la fertilidad masculina.

3.1.5. Fortalezas y debilidades de la revisión

Fortalezas

La revisión presenta varias fortalezas metodológicas relevantes. Se desarrolló a partir de un protocolo previamente publicado y registrado en PROSPERO (CRD42021265401), dentro del programa de revisiones sistemáticas promovido por la OMS, y se llevó a cabo siguiendo el marco metodológico del *WHO handbook for guideline development* y las recomendaciones COSTER. La selección de los estudios se realizó mediante criterios de elegibilidad definidos previamente, con cribado independiente por pares tanto de títulos y resúmenes como de los textos completos, resolución de discrepancias mediante consenso y participación de un tercer revisor cuando fue necesario. La extracción de datos fue igualmente comprobada por un segundo revisor, y tanto el riesgo de sesgo como la certeza de la evidencia se evaluaron mediante procedimientos sistemáticos, utilizando respectivamente las herramientas OHAT y OHAT-GRADE.

La búsqueda bibliográfica fue amplia y sensible. Además de MEDLINE, Embase, Web of Science y EMF-Portal, se exploró literatura gris mediante OpenGrey, IEEE Xplore y otras fuentes, se consultaron sitios web de organismos pertinentes y expertos en CEM-RF, y se examinaron tanto las referencias como las citas posteriores de los estudios incluidos. La ausencia de restricciones por fecha o idioma amplió adicionalmente la cobertura potencial de la búsqueda. Estas características están orientadas a reducir la probabilidad de omitir estudios relevantes por depender exclusivamente de la literatura convencional publicada.

Otra fortaleza reside en la definición relativamente estricta de los criterios de elegibilidad aplicados a la evaluación de la exposición y de los desenlaces elegibles. La SAR se estableció como la métrica ideal de exposición, aunque se aceptaron proxies cuando aquella no estaba disponible, y se excluyeron procedimientos de clasificación particularmente imprecisos, como la simple categorización dicotómica basada en haber poseído o no un teléfono móvil.

Asimismo, para los parámetros seminales se exigieron evaluaciones cuantitativas y se excluyeron otros parámetros cuya validez como indicadores diagnósticos de infertilidad no estuviera establecida. Estas decisiones pretendían limitar la incorporación de evidencia basada en medidas de exposición o desenlaces excesivamente inespecíficos.

En la síntesis cuantitativa, los autores evaluaron inicialmente diferentes posibilidades de combinación de los estudios y descartaron aquellas que consideraron metodológicamente poco justificables. En particular, no realizaron metaanálisis mediante SMD al no poder asumir que las diferencias entre estudios se debieran exclusivamente a las escalas de medida, ni combinaron indiscriminadamente contrastes de exposición alta frente a baja o nula cuando las categorías de exposición eran demasiado heterogéneas. Cuando los datos lo permitieron, optaron por metaanálisis dosis-respuesta mediante modelos lineales de efectos aleatorios, tras comparar previamente el ajuste de modelos lineales y cuadráticos. Esta decisión permitió expresar los resultados mediante cambios en los parámetros seminales por incremento de una hora diaria de uso del teléfono móvil, evitando imponer equivalencias entre categorías de exposición definidas de manera diferente en los estudios originales.

Finalmente, la revisión hace explícitas las desviaciones respecto al protocolo y las restricciones analíticas derivadas de la evidencia disponible. Los autores especifican, entre otras cuestiones, por qué no pudieron realizar análisis de subgrupos, análisis de sensibilidad ni evaluaciones estadísticas del sesgo de publicación, y justifican por qué recurrieron a una síntesis narrativa siguiendo SWiM cuando el metaanálisis no resultaba apropiado. Esta transparencia permite diferenciar las decisiones metodológicas de la revisión de las limitaciones impuestas por los estudios primarios.

Limitaciones

La principal limitación de la revisión procede de la escasez y de las deficiencias metodológicas de la evidencia primaria disponible. Únicamente se identificaron nueve estudios, siete correspondientes a población general y dos a exposiciones ocupacionales. Además, la evidencia estuvo fuertemente concentrada en la exposición localizada asociada al uso de teléfonos móviles, mientras que solo un estudio de población general evaluó otras fuentes de CEM-RF y únicamente dos estudios analizaron exposiciones ocupacionales, uno en técnicos expuestos a microondas y otro en marinos expuestos a radar. Las diferencias entre estos dos últimos estudios en cuanto a población y fuente de exposición hicieron que los autores considerasen inapropiado combinar cuantitativamente sus resultados.

Una limitación especialmente importante afecta a la caracterización de la exposición. Aunque la SAR en los órganos reproductores constituía la métrica de exposición ideal planteada por la revisión, los estudios disponibles recurrieron fundamentalmente a proxies, en particular duración del uso del teléfono móvil, tiempo de conversación o posición del dispositivo respecto al cuerpo. Estas métricas no cuantifican directamente la exposición o la tasa de absorción específica en los órganos reproductores y fueron evaluadas de forma heterogénea entre los estudios. Esta circunstancia contribuyó tanto al riesgo de sesgo en la caracterización de la exposición como a la indirectitud de la evidencia y constituye uno de los principales obstáculos para establecer relaciones exposición-respuesta fiables.

A ello se añade que las características e intensidad de las emisiones de las tecnologías de telefonía móvil han cambiado con el tiempo, lo que dificulta todavía más interpretar una métrica genérica como "tiempo de uso" como una medida homogénea de exposición entre estudios y periodos.

La calidad metodológica de los estudios primarios fue asimismo limitada. Los autores identificaron problemas frecuentes relacionados con la caracterización de la exposición, la selección de los participantes, la evaluación de los desenlaces y el control de la confusión. En los estudios ocupacionales las limitaciones fueron particularmente acusadas: uno presentaba riesgo probablemente o definitivamente alto en todos los dominios evaluados y el otro mostraba riesgo elevado en caracterización de la exposición, evaluación del desenlace y confusión. Por ello, incluso cuando algunos estudios individuales mostraban diferencias estadísticamente significativas en parámetros seminales, su interpretación causal estaba fuertemente condicionada por estas limitaciones.

También existió una heterogeneidad importante en algunos de los metaanálisis dosis-respuesta. Para morfología espermática, I^2 alcanzó el 83,3 %, y para recuento total de espermatozoides fue del 90 %. Aunque la heterogeneidad fue menor o inexistente para otros desenlaces, el reducido número de estudios que contribuyeron a cada análisis, entre dos y tres, limita considerablemente la capacidad para determinar sus causas. Precisamente por este escaso número de estudios no pudieron realizarse análisis de subgrupos ni análisis de sensibilidad destinados a comprobar la robustez de los resultados frente a los supuestos adoptados durante la revisión o frente al riesgo de sesgo de los estudios primarios.

La heterogeneidad en la definición de las categorías de exposición impidió además realizar los metaanálisis tradicionales de exposición alta frente a baja o nula previstos inicialmente. Los autores recurrieron a análisis dosis-respuesta cuando existían datos suficientes y comparables, una decisión metodológicamente justificada, pero que restringió la síntesis cuantitativa a un número reducido de estudios y desenlaces. Del mismo modo, las diferencias entre las fuentes de exposición impidieron realizar los análisis de subgrupos previstos para explorar si los resultados variaban según el origen de los CEM-RF.

Otra limitación fue la imposibilidad de evaluar formalmente el sesgo de publicación o de reporte. Aunque el protocolo contemplaba procedimientos estadísticos específicos, entre ellos la prueba de Egger y otros métodos adaptados al tipo de variable analizada, la cantidad de datos disponible fue insuficiente para aplicarlos. Por tanto, no puede determinarse hasta qué punto la evidencia publicada representa el conjunto completo de los estudios realizados sobre esta cuestión. Tampoco se realizaron análisis de sensibilidad para examinar el efecto de las decisiones adoptadas durante el proceso de revisión o del riesgo de sesgo sobre las estimaciones obtenidas.

En conjunto, la revisión presenta un procedimiento sistemático, transparente y metodológicamente riguroso, pero su capacidad para establecer conclusiones sólidas está condicionada fundamentalmente por la evidencia primaria disponible. El pequeño número de estudios, el predominio casi exclusivo de proxies de exposición relacionados con el uso del teléfono móvil, la ausencia de estimaciones directas de dosis relevantes para los órganos reproductores, el riesgo de sesgo de los estudios y la heterogeneidad de determinadas estimaciones impiden alcanzar conclusiones firmes. Estas limitaciones explican que la certeza de la evidencia para los principales parámetros seminales evaluados se calificase como muy baja y que los autores considerasen muy incierto el posible efecto de la exposición a CEM-RF sobre la fertilidad masculina.

3.1.6. Conclusiones

En conjunto, Kenny et al. concluyen que la evidencia disponible es muy incierta respecto a los posibles efectos de la exposición a CEM-RF sobre la fertilidad masculina. La mayor parte de la evidencia identificada procede de estudios de población general centrados en exposiciones localizadas asociadas al uso del teléfono móvil, mientras que son escasos los trabajos que evalúan otras fuentes de CEM-RF o exposiciones ocupacionales. En los desenlaces para los que fue posible realizar metaanálisis dosis-respuesta, concretamente concentración espermática, morfología, motilidad progresiva y recuento total de espermatozoides, la certeza de la evidencia fue calificada como muy baja mediante OHAT-GRADE.

Los análisis dosis-respuesta no proporcionaron evidencia concluyente de que un mayor tiempo diario de uso del teléfono móvil se asociara con cambios en estos parámetros seminales. Las estimaciones fueron compatibles con ausencia de efecto y presentaron una incertidumbre considerable, a lo que se añade que la mayoría de los estudios incluidos presentaban riesgo de sesgo y que cada desenlace estaba sustentado por un número reducido de estudios. Para la posición del teléfono respecto al cuerpo, los datos disponibles sugieren que llevarlo en el bolsillo delantero puede producir poca o ninguna diferencia en la concentración espermática, el recuento total, la morfología, la motilidad progresiva o el tiempo hasta el embarazo, aunque también en este caso la base empírica es limitada y está afectada por problemas metodológicos. Un único estudio evaluó otras fuentes de CEM-RF en población general y sugirió poca o ninguna diferencia en concentración, motilidad total o recuento espermático asociada al uso de ordenadores u otros dispositivos eléctricos.

La evidencia ocupacional es todavía más limitada. Solo se identificaron dos estudios, uno en técnicos expuestos a microondas y otro en marinos expuestos a equipos de radar, cuyas diferencias en población y fuente de exposición impidieron combinar cuantitativamente los resultados. Además, ambos presentaban problemas importantes de riesgo de sesgo, por lo que estos datos tampoco permiten establecer conclusiones fiables sobre los posibles efectos de las exposiciones ocupacionales a CEM-RF sobre la fertilidad masculina.

Por tanto, la conclusión de la revisión no es que se haya demostrado la ausencia de efectos, sino que la evidencia disponible no permite determinar con confianza si la exposición a CEM-RF afecta a la fertilidad masculina. Los propios autores resumen esta situación señalando que, dado el nivel muy bajo de certeza de los análisis dosis-respuesta, el riesgo de sesgo de la mayoría de los estudios y el reducido número de trabajos disponibles para cada desenlace, no puede confiarse en que el cuerpo de evidencia actual proporcione una respuesta concluyente sobre los efectos de los CEM-RF en la fertilidad masculina.

Implicaciones para investigación futura

Los autores consideran necesaria una investigación prospectiva de mayor calidad y rigor metodológico, tanto en población general como en contextos ocupacionales. Una prioridad fundamental es mejorar la caracterización de la exposición. La evidencia disponible se concentra principalmente en el uso de teléfonos móviles, por lo que se propone ampliar la investigación a un abanico mayor de tecnologías y fuentes de CEM-RF, incluidas exposiciones de campo lejano, como las procedentes de estaciones base.

Las métricas de exposición deberían describirse con precisión, emplear unidades claramente definidas y, siempre que sea posible, cuantificarse objetivamente. Los autores destacan además que la intensidad de los CEM-RF emitidos por los teléfonos móviles ha cambiado con el tiempo, circunstancia que debería tenerse en cuenta en la evaluación y comunicación de la exposición.

Cuando se utilicen proxies basados en el uso del teléfono, estos deberían definirse de forma mucho más precisa. Una medida genérica de «tiempo de uso» puede representar actividades muy diferentes desde el punto de vista de la exposición, por lo que debería especificarse qué incluye exactamente, por ejemplo, tiempo de conversación o utilización de aplicaciones. Además, los estudios deberían considerar explícitamente cómo queda expuesto el aparato reproductor en función del equipo estudiado y valorar la plausibilidad biológica de un posible efecto teniendo en cuenta la distancia a la fuente, la duración y la intensidad de la exposición.

En el ámbito ocupacional, Kenny et al. señalan la necesidad de realizar más estudios prospectivos que incluyan diferentes categorías profesionales y permitan comparar distintos niveles y duraciones de exposición a CEM-RF. Esta recomendación responde a la extrema escasez y heterogeneidad de la evidencia ocupacional actualmente disponible, que impide determinar con fiabilidad los posibles efectos sobre los parámetros reproductivos masculinos.

Los autores proponen asimismo mejorar la homogeneidad de los desenlaces mediante conjuntos estandarizados de resultados, haciendo referencia al *core outcome set* para investigación en infertilidad y al conjunto específico emergente para fertilidad masculina desarrollados dentro de la iniciativa COMMIT (*Core Outcome Measure for Infertility Trials*). Esta estandarización permitiría mejorar la consistencia en la selección, recogida y comunicación de los desenlaces y facilitaría la comparación y síntesis de futuros estudios.

Finalmente, los futuros trabajos deberían recoger y comunicar de forma transparente la información demográfica, de exposición y de análisis disponible, y controlar adecuadamente los potenciales factores de confusión. Los autores mencionan expresamente variables como la edad y otras exposiciones ambientales, que deberían incorporarse en análisis multivariados. En síntesis, Kenny et al. consideran que estudios prospectivos metodológicamente más rigurosos, especialmente con una medición más precisa de la exposición y métodos estadísticos apropiados, son necesarios para ampliar la evidencia disponible y poder alcanzar una mayor certeza sobre los posibles efectos de las exposiciones localizadas y de cuerpo entero a CEM-RF en la salud reproductiva masculina.

3.2. Resultados reproductivos femeninos: estudios observacionales en humanos

3.2.1. Referencia completa del artículo

Título: The effects of radiofrequency exposure on adverse female reproductive outcomes: A systematic review of human observational studies with dose-response meta-analysis. Johnson E.E., Kenny R.P.W., Adesanya A.M., Richmond C., Beyer F., Calderon C., Rankin J., Pearce M.S., Toledano M., Craig D., Pearson F. **Revista y datos bibliográficos:** Environment International, 190, 108816 (2024). DOI: 10.1016/j.envint.2024.108816. **Registro de protocolo:** PROSPERO CRD42021265401.

3.2.2. Objetivos y preguntas de investigación

El objetivo principal de esta revisión sistemática fue evaluar y sintetizar la evidencia procedente de estudios observacionales en humanos sobre los posibles efectos de la exposición a CEM-RF en resultados reproductivos adversos relacionados con el embarazo y el nacimiento.

La pregunta de investigación se formuló mediante el marco Población, Exposición, Comparador, Resultado (PECO). La población de interés comprendió adultos de ambos sexos en el periodo preconcepcional y mujeres embarazadas; la exposición correspondió a CEM-RF procedentes de fuentes ambientales o de uso personal y de exposiciones ocupacionales; el comparador estuvo constituido por personas sin exposición o con niveles inferiores de exposición; y los desenlaces fueron parto prematuro, pequeño para la edad gestacional (SGA), aborto espontáneo, mortinatalidad, bajo peso al nacer (LBW) y anomalías congénitas. De este modo, la revisión abordó tanto exposiciones de la población general como exposiciones específicas del entorno laboral y consideró la exposición relevante durante la preconcepción o el embarazo.

Además de estimar la asociación entre la exposición a CEM-RF y estos desenlaces mediante comparaciones entre los niveles más altos y más bajos de exposición cuando los estudios eran suficientemente comparables, los autores se propusieron examinar posibles relaciones dosis-respuesta cuando los datos disponibles lo permitieran. La revisión combinó, por tanto, una evaluación de las asociaciones entre categorías de exposición con el análisis de posibles gradientes de efecto en función del nivel de exposición, recurriendo a síntesis narrativa cuando la heterogeneidad de los estudios impedía una integración cuantitativa adecuada.

3.2.3. Metodología

Se desarrolló y siguió un protocolo detallado (Kenny et al., 2022) y registrado en PROSPERO (CRD42021266268) y correspondiente a SR3B. El protocolo fue desarrollado conjuntamente para las revisiones observacionales sobre fertilidad masculina (SR3A) y resultados reproductivos femeninos (SR3B), aunque el presente artículo aborda exclusivamente la segunda.

La planificación metodológica siguió el *WHO Handbook for Guideline Development* y las recomendaciones COSTER, y el protocolo se elaboró conforme a PRISMA-P.

Búsqueda y fuentes

La búsqueda bibliográfica se diseñó con un enfoque amplio y sensible. Las principales fuentes fueron MEDLINE, Embase y EMF-Portal. La estrategia se desarrolló iterativamente a partir de los conceptos centrales de la pregunta de investigación, combinando términos libres y encabezamientos temáticos e incorporando desenlaces identificados con la participación de expertos clínicos. No se aplicaron restricciones por fecha ni por idioma. La búsqueda inicial se realizó en noviembre de 2020 y se actualizó en diciembre de 2021 y enero de 2023. Los resultados se gestionaron mediante EndNote, donde se eliminaron los duplicados antes del cribado.

La búsqueda se complementó con literatura gris. En abril de 2021 se consultó OpenGrey y se realizaron búsquedas de guías e informes de organismos de salud pública y protección radiológica, tesis y comunicaciones científicas. También se utilizaron Web of Science, para resúmenes de congresos, e IEEE Xplore®. Estas búsquedas se actualizaron en enero de 2022 y enero de 2023 y, tras dejar de estar disponible OpenGrey, se recurrió a búsquedas avanzadas en Google. Además, se examinaron las referencias de revisiones sistemáticas y narrativas relevantes, las referencias y citas posteriores de los estudios incluidos y trabajos adicionales identificados por expertos en CEM-RF. La actualización final de enero de 2023 aportó un único estudio adicional.

Criterios de elegibilidad

Los criterios de elegibilidad se establecieron a priori mediante el marco PECO. La población de interés comprendió adultos de ambos sexos durante el periodo periconcepcional, tanto materno como paterno, y mujeres embarazadas expuestas a CEM-RF, siempre que se evaluarán resultados adversos del embarazo. También podían incluirse mujeres ya en el periodo posparto y su descendencia cuando la exposición de interés hubiese tenido lugar durante la preconcepción o el embarazo.

Se consideraron tanto exposiciones localizadas como de cuerpo entero. La tasa de absorción específica (SAR), expresada en W/kg, se estableció como la métrica ideal de exposición. Sin embargo, dada la dificultad de disponer de estimaciones de SAR en los órganos reproductores en estudios epidemiológicos, se admitieron también medidas indirectas o *proxies* basadas en niveles medidos o modelizados de campo eléctrico o magnético, densidad de potencia y otras aproximaciones comparables. Para la telefonía móvil se aceptaron medidas autoinformadas, como duración o frecuencia de uso, y medidas objetivas. Para estaciones base se exigieron estimaciones objetivas de la exposición o de la distancia a la fuente, excluyéndose las basadas exclusivamente en distancia autoinformada. También se contemplaron mediciones puntuales, exposímetros personales y modelos de predicción de la exposición.

En el ámbito ocupacional se consideraron las diferentes fuentes potenciales de CEM-RF presentes en el entorno laboral y se aceptaron estimaciones de exposición basadas en mediciones, observación, valoración por expertos, combinaciones de estos procedimientos o matrices empleo-exposición (*Job-Exposure Matrix*, JEM).

Se excluyeron las clasificaciones basadas exclusivamente en el título o categoría profesional, así como medidas excesivamente imprecisas de exposición, por ejemplo, la mera posesión alguna vez de un teléfono móvil. También se excluyeron las exposiciones procedentes de tecnologías médicas cuando los sujetos estudiados eran pacientes, en lugar de trabajadores que utilizaban habitualmente esos equipos.

Los comparadores podían corresponder a ausencia o bajo nivel de exposición frente a niveles superiores, a dos o más categorías de exposición o duración, o a datos continuos que permitieran evaluar una relación dosis-respuesta. Los desenlaces de interés fueron parto prematuro, pequeño para la edad gestacional (SGA), aborto espontáneo, mortinatalidad, bajo peso al nacer (LBW) y anomalías congénitas.

Fueron elegibles estudios observacionales cuantitativos, incluidos estudios de cohortes, casos y controles y estudios transversales que proporcionaran información adecuada sobre la asociación entre exposición y desenlace. Se excluyeron los informes de casos, los estudios con autoselección de participantes a partir de una población de origen no definida y los estudios preclínicos o *in vitro*. No se establecieron restricciones por año o idioma y se contempló tanto literatura publicada como no publicada.

Selección y extracción de datos

La selección se realizó en dos fases, primero mediante cribado de títulos y resúmenes y posteriormente mediante evaluación de los textos completos potencialmente elegibles. Pares de revisores realizaron las evaluaciones de forma independiente y cegada respecto a la decisión del otro revisor. Las discrepancias se resolvieron mediante discusión o, cuando fue necesario, mediante la intervención de un tercer revisor. El procedimiento de cribado fue previamente pilotado entre los distintos pares de revisores para reducir posibles inconsistencias en la aplicación de los criterios de elegibilidad.

La extracción de datos se realizó mediante formularios estandarizados y previamente pilotados. Un revisor efectuó la extracción inicial y un segundo comprobó posteriormente su integridad y exactitud frente a las publicaciones originales. Las discrepancias se resolvieron mediante discusión y, cuando fue necesario, con la intervención de un tercer revisor.

Medidas de efecto y síntesis cuantitativa

La estrategia analítica contempló tanto metaanálisis pareados, comparando los niveles más altos y más bajos de exposición, como metaanálisis dosis-respuesta cuando los datos disponibles permitían realizarlos de forma apropiada y plausible. Los metaanálisis pareados se realizaron mediante modelos de efectos aleatorios y se utilizaron fundamentalmente medidas relativas de asociación, principalmente RR u OR según el diseño y la información proporcionada por los estudios. La heterogeneidad entre estudios se evaluó mediante I^2 y la estimación de la varianza entre estudios, y se calcularon intervalos de predicción cuando procedía.

Los análisis dosis-respuesta se realizaron cuando al menos dos estudios proporcionaban información suficientemente comparable sobre un mismo desenlace y una medida cuantificable de exposición.

Para ello se utilizó el paquete *dosresmeta* en R. Se exploraron modelos lineales y cuadráticos no lineales y su ajuste se comparó mediante pruebas de bondad de ajuste y criterios de información. Cuando los estudios presentaban categorías de exposición, se asignaron valores continuos representativos a dichas categorías para posibilitar su incorporación a los modelos dosis-respuesta.

A diferencia de la revisión de fertilidad masculina, en Johnson et al. los metaanálisis pareados constituyeron una parte central de la síntesis cuantitativa. En población general se aplicaron a parto prematuro, bajo peso al nacer y SGA, mientras que en los estudios ocupacionales se utilizaron, cuando fue posible, para aborto espontáneo, parto prematuro y bajo peso al nacer. Los análisis dosis-respuesta tuvieron especial relevancia para determinados desenlaces ocupacionales, en particular aborto espontáneo y anomalías congénitas en fisioterapeutas expuestas a diatermia de onda corta.

Cuando los datos no permitieron una combinación cuantitativa adecuada, los resultados se sintetizaron narrativamente siguiendo las recomendaciones SWiM (*Synthesis Without Meta-analysis*). Esto afectó, entre otros, a determinados análisis de aborto espontáneo y anomalías congénitas en población general y a SGA y mortinatalidad en estudios ocupacionales. El reducido número de estudios disponibles limitó asimismo la realización de determinados análisis adicionales.

Evaluación del riesgo de sesgo (RoB)

El riesgo de sesgo se evaluó mediante la herramienta OHAT del National Toxicology Program, aplicada a nivel de estudio y desenlace. Se consideraron siete dominios: selección o participación, caracterización de la exposición, evaluación del desenlace, confusión, pérdidas o exclusiones, notificación selectiva y otras amenazas potenciales a la validez interna. Cada dominio se clasificó como riesgo definitivamente bajo, probablemente bajo, probablemente alto o definitivamente alto.

La evaluación fue realizada por dos revisores, resolviéndose las discrepancias mediante consenso y, cuando fue necesario, mediante la participación de un tercer revisor. Para el dominio de confusión se identificaron a priori factores considerados críticos o relevantes para la relación entre exposición y resultados reproductivos. La ausencia de un control adecuado de la confusión no determinaba automáticamente la exclusión de un estudio, pero sí se incorporaba al juicio sobre su riesgo de sesgo. La evaluación consideró también la calidad y precisión de la caracterización de la exposición, un aspecto especialmente relevante dada la heterogeneidad de las métricas empleadas en los estudios incluidos. De hecho, la propia revisión destaca posteriormente que la medición de la exposición fue con frecuencia inconsistente y estuvo insuficientemente descrita.

Certeza de la evidencia (OHAT-GRADE)

La certeza de la evidencia se evaluó mediante OHAT-GRADE para los conjuntos de estudios que pudieron sintetizarse cuantitativamente. Se consideraron los dominios de riesgo de sesgo, inconsistencia, indirectitud, imprecisión y sesgo de publicación, junto con los factores que podían justificar una mejora de la certeza. El nivel inicial dependió del diseño de los estudios, partiendo las cohortes de un nivel de certeza moderado, y posteriormente se aplicaron las correspondientes degradaciones o mejoras de acuerdo con las características del cuerpo de evidencia.

La valoración se realizó por desenlace y contexto de exposición, distinguiendo cuando correspondía entre población general y población ocupacional. Dos revisores participaron en la evaluación de la certeza y las discrepancias se resolvieron mediante consenso o con la intervención de un tercer revisor. Las conclusiones se expresaron mediante la terminología asociada a los niveles de certeza de OHAT-GRADE. Cuando los estudios únicamente pudieron sintetizarse narrativamente, no se realizó una evaluación completa de certeza mediante OHAT-GRADE y la interpretación se apoyó fundamentalmente en el riesgo de sesgo de los estudios disponibles.

Una cuestión metodológica especialmente relevante fue la indirectitud de la exposición. En los estudios de población general, gran parte de la evidencia se basó en *proxies*, principalmente tiempo de utilización o de llamada con teléfonos móviles, en lugar de estimaciones dosimétricas directas. En determinados análisis ocupacionales, la indirectitud estuvo relacionada con las fuentes específicas y heterogéneas de exposición y, para las anomalías congénitas, también con la agrupación de malformaciones de naturaleza diversa en un único desenlace.

3.2.4. Resultados

Flujo PRISMA

La búsqueda y el proceso inicial de selección se realizaron conjuntamente para las dos revisiones observacionales sobre reproducción, SR3A, centrada en fertilidad masculina, y SR3B, sobre resultados reproductivos femeninos. Tras la eliminación de duplicados, se cribaron 20.329 registros mediante título y resumen. De ellos, 278 informes fueron seleccionados para su recuperación en texto completo, de los cuales nueve no pudieron obtenerse. El encadenamiento de citas y la identificación de estudios por expertos aportaron otros 43 registros que fueron evaluados adicionalmente.

Finalmente, 18 estudios fueron incluidos en la revisión de resultados reproductivos femeninos: ocho correspondían a exposiciones en población general y diez a exposiciones ocupacionales. Tres publicaciones evaluaban poblaciones solapadas de miembros varones de la Marina Real Noruega; para evitar contabilizar repetidamente a los mismos participantes, los autores conservaron para la síntesis únicamente Baste et al. (2012), por ser el estudio más reciente y de mayor tamaño.

Características de los estudios incluidos

Los ocho estudios realizados en población general fueron Baste et al. (2015), Boileau et al. (2020), Brizzi y Marinelli (2018), Karuserci et al. (2019), Lu et al. (2017), Mahmoudabadi et al. (2015), Tsarna et al. (2019) y Zhao et al. (2021). Los diez estudios ocupacionales fueron Allam (2016), Baste et al. (2012), Cromie et al. (2002), Källén et al. (1982), Khan et al. (2018), Kolmodin-Hedman et al. (1988), Lerman et al. (2001), Ouellet-Hellstrom y Stewart (1993), Taskinen et al. (1990) y Xu et al. (2016).

La evidencia procedente de población general se concentró principalmente en la exposición asociada al uso del teléfono móvil, aunque también se estudiaron otras fuentes.

Las métricas de exposición fueron heterogéneas e incluyeron duración del uso del teléfono, tiempo de llamada, número de llamadas, utilización de diferentes dispositivos y, en algunos trabajos, estimaciones o mediciones ambientales de exposición. Brizzi y Marinelli evaluaron la exposición residencial relacionada con radar, mientras que Mahmoudabadi et al. utilizaron una estimación de SAR efectiva asociada al uso del teléfono móvil.

Tsarna et al. (2019) tuvo especial peso en los análisis cuantitativos porque integró datos procedentes de cuatro cohortes de nacimiento independientes. Por este motivo, algunos metaanálisis incluyeron un número de cohortes superior al número de publicaciones. Baste et al. (2015), por su parte, proporcionó información de una cohorte de gran tamaño y analizó tanto exposición materna durante el embarazo como exposición paterna previa a la concepción.

Los estudios ocupacionales fueron considerablemente más heterogéneos respecto a las fuentes y circunstancias de exposición. Incluyeron principalmente fisioterapeutas expuestas a diatermia de onda corta o de microondas, trabajadoras expuestas a equipos de soldadura de plásticos, cajeras potencialmente expuestas a sistemas electrónicos antirrobo y personal militar expuesto a sistemas de radar y comunicaciones. La exposición se estimó mediante horas o frecuencia de utilización de los equipos, distancia a las fuentes, mediciones del campo eléctrico o estimaciones individuales de exposición.

La información sobre localización corporal de la exposición, distribución, niveles de fondo, intensidad y posibles coexposiciones fue incompleta o se notificó de forma heterogénea. Esta diversidad en las fuentes y métricas de exposición fue una característica importante del conjunto de evidencia y limitó la posibilidad de combinar cuantitativamente numerosos resultados.

Los desenlaces evaluados comprendieron parto prematuro, pequeño para la edad gestacional (SGA), bajo peso al nacer (LBW), aborto espontáneo, mortinatalidad y anomalías congénitas. No todos los estudios analizaron todos los desenlaces y, para algunos de ellos, el número de investigaciones disponibles fue muy reducido.

Resultados cuantitativos en población general

En población general fue posible realizar metaanálisis pareados para tres desenlaces: parto prematuro, SGA y bajo peso al nacer. Los análisis compararon categorías de exposición media o alta con categorías de exposición nula o baja, fundamentalmente en relación con el uso materno de teléfonos móviles.

Para el parto prematuro se combinaron cuatro estudios. Dado que Tsarna et al. aportó cuatro cohortes independientes, el metaanálisis comprendió en conjunto siete cohortes y 153.929 participantes. El riesgo relativo agrupado fue 1,14 (IC95 % 0,97–1,34), con un intervalo de predicción del 95 % de 0,83–1,57. Por tanto, la estimación puntual sugería un pequeño incremento del riesgo, pero el intervalo de confianza incluía la ausencia de asociación y el intervalo de predicción era compatible con efectos en ambas direcciones. La certeza de la evidencia se calificó como muy baja.

Para SGA se combinaron dos estudios, aunque la inclusión de las cuatro cohortes independientes de Tsarna et al. hizo que el análisis comprendiera cinco cohortes y 153.027 participantes. El riesgo relativo agrupado fue 1,13 (IC95 % 1,02–1,24), con un intervalo de predicción del 95 % de 0,99–1,28.

Aunque el intervalo de confianza de la estimación agrupada no incluía la unidad, el intervalo de predicción sí lo hacía y la certeza de la evidencia fue muy baja. Los autores consideran, por tanto, que la evidencia es muy incierta respecto al efecto de la exposición materna a CEM-RF sobre el riesgo de SGA.

Para el bajo peso al nacer se incluyeron cuatro estudios, nuevamente equivalentes a siete cohortes por la contribución de Tsarna et al., con 153.929 participantes. El riesgo relativo agrupado fue 1,14 (IC95 % 0,96–1,36), con un intervalo de predicción del 95 % de 0,84–1,57. El resultado fue compatible con ausencia de asociación y la certeza se calificó como muy baja. Los análisis ajustados presentados por Baste et al. y Tsarna et al. tampoco mostraron asociaciones estadísticamente significativas entre las categorías de exposición materna al teléfono móvil y el bajo peso al nacer.

En conjunto, los tres metaanálisis de población general no proporcionaron evidencia de certeza suficiente para establecer un incremento del riesgo de parto prematuro, SGA o bajo peso al nacer asociado a la exposición materna a CEM-RF. Aunque la estimación agrupada para SGA fue superior a la unidad y su IC95 % no incluyó el valor nulo, la muy baja certeza de la evidencia y el intervalo de predicción compatible con ausencia de efecto impiden interpretar este resultado como evidencia robusta de asociación.

Aborto espontáneo y anomalías congénitas en población general

No fue posible realizar metaanálisis para aborto espontáneo ni anomalías congénitas en población general debido a la heterogeneidad de las exposiciones y de los desenlaces estudiados.

Dos estudios evaluaron el aborto espontáneo, pero no pudieron combinarse debido a las diferencias sustanciales en la caracterización de la exposición. Brizzi y Marinelli (2018), que estudiaron exposición ambiental relacionada con radar, encontraron poca o ninguna diferencia global en el riesgo de aborto espontáneo entre la población expuesta y la no expuesta, aunque las estimaciones variaron entre las distintas categorías de exposición. Mahmoudabadi et al. (2015), por el contrario, informó una asociación entre una mayor estimación de exposición asociada al uso del teléfono móvil, basada en una estimación de SAR efectiva, y el aborto espontáneo. Los resultados fueron, por tanto, discordantes y ambos estudios presentaban limitaciones metodológicas y riesgo de sesgo, por lo que los autores consideran incierto el efecto de la exposición a CEM-RF sobre este desenlace.

Cuatro estudios evaluaron anomalías congénitas: Baste et al. (2015), Boileau et al. (2020), Brizzi y Marinelli (2018) y Zhao et al. (2021). Tampoco pudieron combinarse cuantitativamente debido a la heterogeneidad de las anomalías consideradas.

Baste et al. no encontró diferencias apreciables en el riesgo global de anomalías congénitas asociadas a exposición materna al teléfono móvil entre las semanas 15 y 30 de gestación: los RR ajustados fueron 0,99 (IC95 % 0,92–1,06) para exposición media y 1,01 (IC95 % 0,92–1,11) para exposición alta, frente a exposición baja. Tampoco se observó una asociación clara con la exposición paterna previa a la concepción.

Zhao et al. (2021), por el contrario, evaluó específicamente cardiopatías congénitas y comunicó estimaciones crecientes con una mayor duración diaria de utilización materna del teléfono móvil. Durante los tres meses anteriores al embarazo, frente a las mujeres no expuestas, los OR fueron 1,17 (IC95 % 0,75–1,84) para menos de 180 minutos diarios, 1,35 (0,83–2,19) para 180–300 minutos y 2,70 (1,69–4,30) para más de 300 minutos diarios.

Durante el primer trimestre se observó un patrón semejante, con OR de 1,23 (0,80–1,89), 1,25 (0,78–1,99) y 1,80 (1,14–2,83), respectivamente. Estos resultados proceden de un único estudio y no pudieron integrarse con los restantes trabajos sobre anomalías congénitas, por lo que no permiten establecer por sí solos una asociación general.

Los otros dos estudios, Boileau et al. (2020) y Brizzi y Marinelli (2018), evaluaron asimismo anomalías congénitas, pero emplearon definiciones y exposiciones diferentes de las utilizadas por Baste et al. y Zhao et al. Esta heterogeneidad, tanto en el tipo de anomalía considerada como en la caracterización de la exposición, impidió integrar cuantitativamente los cuatro estudios. En consecuencia, los resultados individuales variaron entre trabajos y no proporcionaron un patrón consistente que permitiera establecer una asociación entre exposición a CEM-RF y anomalías congénitas.

Ningún estudio realizado en población general evaluó la asociación entre exposición a CEM-RF y mortinatalidad.

Resultados en estudios ocupacionales

En los estudios ocupacionales fue posible realizar tanto metaanálisis dosis-respuesta como metaanálisis pareados para determinados desenlaces. La diversidad de las fuentes de exposición fue considerablemente mayor que en población general e incluyó diatermia de onda corta y de microondas, equipos de soldadura, sistemas electrónicos de vigilancia y radar.

Para aborto espontáneo se realizó, en primer lugar, un metaanálisis dosis-respuesta de dos estudios de casos y controles, Lerman et al. (2001) y Taskinen et al. (1990), ambos en fisioterapeutas expuestas a diatermia de onda corta. Por cada hora de utilización, la OR estimada fue 1,02 (IC95 % 0,94–1,10), con $I^2 = 72\%$ y $\tau = 0,017$. La estimación fue muy próxima a la ausencia de efecto y la certeza de la evidencia fue muy baja.

Además, cinco estudios con diferentes fuentes de exposición pudieron combinarse mediante un metaanálisis pareado. El RR agrupado de aborto espontáneo en trabajadoras expuestas frente a no expuestas fue 1,06 (IC95 % 0,96–1,18). También en este caso la certeza fue muy baja y la estimación fue compatible con poca o ninguna diferencia entre grupos.

Para anomalías congénitas, Lerman et al. y Taskinen et al. permitieron realizar un análisis dosis-respuesta de la exposición a diatermia de onda corta en fisioterapeutas. La prueba de no linealidad indicó que el modelo lineal no era adecuado ($\chi^2 = 7,47$; $gl = 1$; $p = 0,006$), por lo que los autores utilizaron un modelo cuadrático. La estimación presentada fue OR 1,40 (IC95 % 0,85–2,32), con $I^2 = 68,1\%$ y $\tau = 0,31$. La certeza fue muy baja y el intervalo de confianza fue compatible tanto con ausencia de asociación como con un incremento del riesgo.

Otros estudios ocupacionales sobre anomalías congénitas no pudieron incorporarse a este análisis. Baste et al. (2012), en personal masculino de la Marina Real Noruega, presentó resultados muy imprecisos y sin un patrón exposición-respuesta consistente según la intensidad y proximidad temporal de la exposición. Kolmodin-Hedman et al. (1988) comunicó cuatro anomalías menores entre trabajadoras de fábricas, pero no especificó en qué grupos de exposición se produjeron ni el tipo concreto de anomalía, lo que impidió interpretar adecuadamente el resultado.

Para parto prematuro se combinaron tres estudios sobre exposición materna ocupacional: Allam (2016), Lerman et al. (2001) y Xu et al. (2016). El metaanálisis de datos no ajustados produjo un RR de 1,19 (IC95 % 0,32–4,37), con un intervalo de predicción del 95 % de 0,18–7,87. La extraordinaria amplitud de estos intervalos refleja una gran imprecisión y la certeza de la evidencia fue muy baja.

Baste et al. (2012) aportó además información sobre parto prematuro asociada a exposición paterna ocupacional en personal de la Marina Real Noruega, que no pudo integrarse con el metaanálisis anterior de exposición materna. Las estimaciones variaron considerablemente según el nivel de exposición y el intervalo transcurrido entre esta y la concepción, con intervalos de confianza muy amplios y sin un patrón exposición-respuesta consistente. Estos resultados adicionales tampoco permiten establecer una asociación clara entre la exposición paterna ocupacional a CEM-RF y el riesgo de parto prematuro.

Para bajo peso al nacer se combinaron los mismos tres estudios que evaluaron la exposición materna ocupacional y el parto prematuro. El metaanálisis mostró un RR de 2,90 (IC95 % 0,69–12,23). Aunque la estimación puntual fue superior a la unidad, el intervalo de confianza fue extremadamente amplio e incluyó la ausencia de asociación, por lo que la evidencia se consideró muy incierta y de certeza muy baja.

Para mortinatalidad, los estudios ocupacionales disponibles no pudieron combinarse mediante metaanálisis. Los acontecimientos fueron muy escasos y los estudios utilizaron definiciones, fuentes de exposición y comparadores diferentes. Cromie et al. (2002), Källén et al. (1982), Ouellet-Hellstrom y Stewart (1993) y Xu et al. (2016) aportaron información sobre este desenlace, pero el reducido número de casos y la heterogeneidad metodológica impidieron obtener una estimación conjunta interpretable. Los resultados fueron sintetizados narrativamente y no mostraron un patrón consistente de incremento del riesgo asociado a la exposición ocupacional a CEM-RF.

Para SGA, tampoco fue posible realizar metaanálisis. Baste et al. (2012) evaluó la asociación con exposición paterna ocupacional a sistemas de radar y comunicaciones, mientras que Khan et al. (2018) estudió exposición materna ocupacional. Las estimaciones de Baste et al. variaron según la intensidad y la proximidad temporal de la exposición respecto a la concepción, con intervalos de confianza muy amplios y sin un gradiente exposición-respuesta consistente. Khan et al. obtuvo asimismo una estimación imprecisa compatible tanto con una reducción como con un incremento del riesgo. En conjunto, los resultados de ambos estudios fueron inconsistentes y no permiten determinar si la exposición ocupacional a CEM-RF afecta al riesgo de SGA.

Resultados cualitativos y consistencia global

La síntesis conjunta muestra una evidencia considerablemente heterogénea. En población general, los análisis cuantitativos de parto prematuro y bajo peso al nacer proporcionaron estimaciones agrupadas cuyos intervalos de confianza incluían la ausencia de asociación. Para SGA, la estimación agrupada fue ligeramente superior a la unidad y su intervalo de confianza no incluyó el nulo, aunque el intervalo de predicción sí lo hizo y la certeza fue muy baja.

Para aborto espontáneo y anomalías congénitas en población general no fue posible realizar metaanálisis. Los resultados individuales variaron entre estudios y, particularmente para anomalías congénitas, las diferencias en las definiciones de los desenlaces impidieron una síntesis cuantitativa adecuada.

En los estudios ocupacionales, los metaanálisis disponibles tampoco proporcionaron estimaciones suficientemente precisas como para establecer efectos de la exposición. El análisis dosis-respuesta de aborto espontáneo en fisioterapeutas produjo una estimación próxima al valor nulo, mientras que el análisis de anomalías congénitas y los metaanálisis de parto prematuro y bajo peso al nacer presentaron una incertidumbre considerable. Para mortinatalidad y SGA los datos fueron insuficientes para una síntesis cuantitativa.

En conjunto, Johnson et al. consideran que la evidencia disponible es muy incierta respecto al efecto de la exposición a CEM-RF sobre los resultados reproductivos femeninos.

Evaluación del riesgo de sesgo

El riesgo de sesgo constituyó una limitación importante del conjunto de evidencia. Los problemas se distribuyeron entre prácticamente todos los dominios evaluados mediante la herramienta OHAT, aunque el uso de métodos estadísticos apropiados fue el dominio que presentó menos problemas. En concreto, diez estudios presentaron problemas relacionados con la identificación y el manejo de factores de confusión y tres mostraron riesgo de notificación selectiva.

La caracterización de la exposición fue una fuente destacada de preocupación. Los estudios utilizaron métricas muy diferentes y, especialmente en población general, predominaron medidas indirectas como duración del uso del teléfono o tiempo de llamada. En los estudios ocupacionales, la exposición se caracterizó mediante horas o frecuencia de utilización de equipos, distancia a las fuentes, mediciones de campo o estimaciones individuales. La información sobre localización corporal, intensidad, exposición de fondo y coexposiciones fue con frecuencia incompleta o heterogénea.

La confusión constituyó otra limitación relevante. Varios estudios no controlaron adecuadamente factores considerados críticos, entre ellos edad, índice de masa corporal, nivel socioeconómico, tabaquismo y consumo de alcohol. También se identificaron problemas en la evaluación de los desenlaces, pérdidas o exclusiones de participantes y, en algunos trabajos, posible notificación selectiva.

En relación con los métodos estadísticos, la mayoría de los estudios fueron considerados adecuados, aunque Karuserci et al. (2019) y Källén et al. (1982) fueron clasificados como inadecuados en este dominio. En conjunto, la frecuencia y distribución de las limitaciones metodológicas redujeron sustancialmente la confianza que podía depositarse en las asociaciones observadas en los estudios individuales.

Certeza de la evidencia (OHAT-GRADE)

La certeza de la evidencia se evaluó mediante OHAT-GRADE para los desenlaces que pudieron someterse a metaanálisis. En todos los metaanálisis principales, la certeza final fue muy baja.

En población general, la evidencia sobre parto prematuro, SGA y bajo peso al nacer fue calificada como de certeza muy baja. Para parto prematuro se degradó la certeza por riesgo de sesgo, inconsistencia e indirectitud. Esta última estuvo relacionada principalmente con el predominio del tiempo de llamada como *proxy* de exposición. Para SGA también se aplicaron degradaciones por riesgo de sesgo, inconsistencia e indirectitud. En el caso del bajo peso al nacer, la certeza se redujo por riesgo de sesgo, inconsistencia e indirectitud asociada al uso de horas de utilización del teléfono como aproximación a la exposición relevante.

En los estudios ocupacionales, tanto el análisis dosis-respuesta como el análisis pareado de aborto espontáneo fueron calificados como de certeza muy baja. Lo mismo ocurrió con las anomalías congénitas, el parto prematuro y el bajo peso al nacer. Las razones variaron entre desenlaces, pero incluyeron riesgo de sesgo, inconsistencia, indirectitud de las métricas o fuentes de exposición e imprecisión. Para anomalías congénitas, la combinación de malformaciones diferentes en una categoría global añadió además indirectitud al desenlace.

No se aplicaron mejoras de certeza en ninguno de los análisis. Para aquellos desenlaces que no pudieron combinarse cuantitativamente, como aborto espontáneo y anomalías congénitas en determinados estudios de población general y mortinatalidad y SGA en estudios ocupacionales, los autores consideraron el riesgo de sesgo del conjunto de estudios en la síntesis narrativa, pero no realizaron una evaluación OHAT-GRADE formal equivalente a la de los metaanálisis.

En conjunto, los resultados de Johnson et al. indican que la mayor parte de las estimaciones disponibles sugieren poca o ninguna diferencia en los resultados reproductivos femeninos asociada a la exposición a CEM-RF, pero la evidencia presenta limitaciones metodológicas importantes y una certeza predominantemente muy baja. Por ello, los autores consideran que no puede depositarse confianza suficiente en el cuerpo actual de evidencia para determinar con precisión el efecto de la exposición a CEM-RF sobre estos desenlaces.

3.2.5. Fortalezas y debilidades de la revisión

Fortalezas

Una fortaleza principal de la revisión es la amplitud del proceso de búsqueda y selección de la evidencia. Los autores realizaron una búsqueda bibliográfica exhaustiva, complementada con encadenamiento de citas hacia adelante y hacia atrás y consulta a expertos en CEM-RF, con el objetivo de reducir la posibilidad de omitir estudios elegibles. El cribado se efectuó en duplicado y de forma cegada, y se pilotó previamente entre los distintos pares de revisores para reducir posibles discrepancias en la aplicación de los criterios de elegibilidad. La revisión se apoyó además en un protocolo previamente publicado y registrado en PROSPERO y se desarrolló dentro del marco metodológico del *WHO Handbook for Guideline Development* y las recomendaciones COSTER.

La definición previa de la pregunta PECO y de los criterios de elegibilidad permitió incluir tanto exposiciones de la población general como ocupacionales y considerar exposición materna y paterna durante periodos relevantes para el embarazo. Asimismo, la revisión contempló diferentes fuentes de CEM-RF y un conjunto amplio de desenlaces reproductivos, incluidos parto prematuro, SGA, bajo peso al nacer, aborto espontáneo, mortinatalidad y anomalías congénitas.

Otra fortaleza reside en el enfoque utilizado para la síntesis cuantitativa. Cuando los estudios eran suficientemente comparables se realizaron metaanálisis pareados, y cuando la información sobre exposición lo permitía se efectuaron además análisis dosis-respuesta. Los autores evitaron combinar estudios cuando las diferencias en exposición o desenlace hacían poco defendible una estimación agrupada y recurrieron en esos casos a síntesis narrativa. Esta estrategia permitió mantener una distinción metodológica entre resultados que podían sintetizarse cuantitativamente y aquellos cuya heterogeneidad impedía una combinación adecuada.

El uso de la herramienta OHAT para evaluar el riesgo de sesgo y de OHAT-GRADE para establecer la certeza de la evidencia constituye otra fortaleza relevante. Esto permitió valorar de manera explícita problemas de confusión, caracterización de la exposición, evaluación del desenlace, inconsistencia, indirectitud e imprecisión, y evitar que la interpretación dependiera únicamente de la significación estadística de resultados individuales.

Finalmente, los autores describen de forma transparente ciertas limitaciones del propio proceso de revisión. Aunque la extracción inicial fue realizada por un único revisor, un segundo revisor comprobó posteriormente la exactitud de los datos. Asimismo, reconocen que el empleo de distintos pares de revisores durante el cribado podía introducir cierta variabilidad en los juicios de elegibilidad, aunque el pilotaje previo y el cribado independiente en duplicado se utilizaron precisamente para reducir ese riesgo.

Limitaciones

La principal limitación de la revisión deriva de la cantidad, heterogeneidad y calidad metodológica de la evidencia primaria disponible. Solo un número reducido de estudios contribuyó a cada desenlace y muchos presentaban riesgo de sesgo en uno o más dominios relevantes. Diez estudios mostraron problemas en la identificación o manejo de factores de confusión y tres presentaron riesgo de notificación selectiva. En conjunto, el control insuficiente de la confusión fue uno de los problemas más importantes del cuerpo de evidencia.

La caracterización de la exposición constituye otra limitación central. En población general, la mayoría de los estudios utilizaron *proxies* como tiempo de uso del teléfono móvil o duración de las llamadas, con diferencias importantes en la forma de definir y medir estas variables. La localización corporal de la exposición, su intensidad, distribución, niveles de fondo y posibles coexposiciones se notificaron de manera incompleta o heterogénea. Estas limitaciones dificultan relacionar directamente las métricas empleadas con la exposición relevante para los órganos reproductores y contribuyen a la indirectitud de la evidencia.

En los estudios ocupacionales, la heterogeneidad de las fuentes fue todavía mayor. La evidencia procedía principalmente de fisioterapeutas expuestas a diatermia de onda corta, junto con estudios sobre soldadura de plásticos, sistemas electrónicos antirrobo, radar y otras fuentes. Esta diversidad, unida a diferencias en la evaluación de la exposición, limitó la comparabilidad entre trabajos. Los propios autores señalan que en futuros estudios ocupacionales sería preferible realizar mediciones directas con equipos adecuados o, al menos, utilizar matrices empleo-exposición.

La heterogeneidad de los desenlaces también fue relevante, especialmente en el caso de las anomalías congénitas. Muchos estudios agruparon bajo esta denominación alteraciones estructurales o funcionales clínicamente muy diferentes, lo que introduce indirectitud y dificulta determinar si determinados tipos específicos de malformación pudieran asociarse de manera distinta con la exposición. Los autores consideran que la falta de una clasificación más granular impide evaluar adecuadamente riesgos para anomalías individuales o subgrupos clínicamente homogéneos.

El reducido número de estudios y la heterogeneidad metodológica limitaron además la posibilidad de realizar análisis adicionales y de interpretar con precisión algunas estimaciones.

En varios metaanálisis los intervalos de confianza fueron amplios, especialmente en los estudios ocupacionales sobre parto prematuro y bajo peso al nacer, lo que refleja una importante imprecisión. En otros desenlaces, como aborto espontáneo y anomalías congénitas en población general o mortinatalidad y SGA en estudios ocupacionales, la evidencia fue demasiado heterogénea o escasa para permitir una síntesis cuantitativa.

También debe considerarse como limitación del propio proceso de revisión que la extracción de datos no se realizó de forma independiente por duplicado. Un único revisor efectuó la extracción y un segundo comprobó posteriormente los datos. Los autores reconocen que este procedimiento puede generar más errores que una extracción duplicada independiente, aunque la verificación posterior se utilizó para reducir ese riesgo. Del mismo modo, el empleo de diferentes pares de revisores durante el cribado podía introducir cierta variabilidad en la interpretación de los criterios de elegibilidad, pese al pilotaje previo.

En conjunto, la revisión presenta una metodología sistemática, explícita y rigurosa, pero sus conclusiones están condicionadas por una evidencia primaria escasa, heterogénea y frecuentemente afectada por riesgo de sesgo. La utilización predominante de medidas indirectas de exposición, la insuficiente evaluación de factores de confusión, la heterogeneidad de las fuentes ocupacionales y de algunos desenlaces, y la imprecisión de determinadas estimaciones explican que los análisis fueran calificados con certeza baja o muy baja y que los autores no puedan establecer con confianza el efecto de los CEM-RF sobre los resultados reproductivos femeninos.

3.2.6. Conclusiones

En conjunto, Johnson et al. concluyen que la mayor parte de la evidencia disponible sugiere poco o ningún efecto de la exposición a CEM-RF sobre los resultados reproductivos femeninos evaluados. Sin embargo, esta interpretación está condicionada por la baja o muy baja certeza de la evidencia, el riesgo de sesgo de numerosos estudios y el reducido número de trabajos disponibles para cada desenlace, por lo que los autores subrayan que no puede depositarse una confianza elevada en lo que indica actualmente el conjunto de la investigación.

En población general, los metaanálisis realizados no mostraron un incremento claro del riesgo relativo de parto prematuro o bajo peso al nacer asociado a una mayor exposición materna a CEM-RF. Para SGA, la estimación agrupada fue ligeramente superior a la unidad, RR 1,13 (IC95 % 1,02–1,24), aunque la certeza fue muy baja y el intervalo de predicción incluyó la ausencia de efecto. Para aborto espontáneo y anomalías congénitas no fue posible realizar metaanálisis debido a la heterogeneidad de las exposiciones y los desenlaces, y los resultados individuales fueron variables. Ningún estudio de población general evaluó mortinatalidad.

En los estudios ocupacionales, los metaanálisis tampoco proporcionaron evidencia suficientemente sólida de un aumento del riesgo. El análisis dosis-respuesta en fisioterapeutas expuestas a diatermia de onda corta mostró una OR de 1,02 (IC95 % 0,94–1,10) para aborto espontáneo, compatible con poca o ninguna diferencia. Para anomalías congénitas, la estimación fue OR 1,40 (IC95 % 0,85–2,32), también con certeza muy baja. Los metaanálisis de parto prematuro y bajo peso al nacer presentaron intervalos de confianza muy amplios, reflejando una elevada imprecisión, mientras que mortinatalidad y SGA no pudieron sintetizarse cuantitativamente.

Por tanto, la conclusión de la revisión no debe interpretarse como una demostración firme de ausencia de efecto. Los autores señalan que, aunque los metaanálisis disponibles tienden a no mostrar un aumento del riesgo relativo de resultados reproductivos adversos asociados a la exposición a CEM-RF, la evidencia es insuficiente para establecer conclusiones con alta confianza. En los estudios que no pudieron metaanalizarse existió variabilidad en la dirección y magnitud de los efectos, y la mayoría de los trabajos incluidos presentaban algún riesgo de sesgo. En consecuencia, la evidencia actual apunta predominantemente hacia poco o ningún efecto, pero sigue existiendo una incertidumbre considerable.

Implicaciones para investigación futura

Johnson et al. consideran prioritario mejorar la definición y estandarización de los desenlaces, especialmente en el caso de las anomalías congénitas. Estas constituyen un grupo clínicamente muy heterogéneo que incluye alteraciones estructurales y funcionales de gravedad diversa, y con frecuencia los estudios agrupan múltiples entidades diferentes bajo una única categoría. Los autores proponen que futuras investigaciones informen las anomalías con un mayor nivel de detalle y mediante categorías clínicamente homogéneas, de forma que pueda evaluarse el riesgo de anomalías individuales o de subgrupos específicos, algo que la evidencia actual no permite hacer.

También consideran necesario ampliar el espectro de tecnologías y fuentes de exposición evaluadas en población general. La mayor parte de la evidencia disponible se centra en teléfonos móviles, por lo que deberían estudiarse con mayor detalle otras fuentes potencialmente relevantes, incluidas exposiciones de campo lejano, como las estaciones base. Paralelamente, es necesario mejorar sustancialmente la caracterización de la exposición, utilizando métricas más detalladas, objetivas y directamente relacionadas con la exposición relevante para los órganos reproductores.

En los estudios ocupacionales, los autores recomiendan incrementar el uso de mediciones directas mediante equipos apropiados o, cuando ello no sea posible, recurrir al menos a matrices empleo-exposición. La evidencia ocupacional disponible está muy concentrada en fisioterapeutas expuestas a diatermia de onda corta, por lo que se necesitan estudios en otros puestos de trabajo y con diferentes fuentes, intensidades y duraciones de exposición para poder valorar si los resultados son generalizables a otros contextos laborales.

Otra prioridad explícita es estudiar con mayor profundidad la exposición paterna a CEM-RF y su posible relación con los resultados del embarazo. Los autores consideran que este aspecto está insuficientemente representado tanto en población general como en contextos ocupacionales y que debería incorporarse de forma sistemática a futuros estudios.

Finalmente, Johnson et al. recomiendan mejorar la transparencia y exhaustividad del reporte. Los futuros estudios deberían recoger y comunicar toda la información disponible sobre características demográficas, exposición y análisis estadístico, lo que facilitaría la comparación y combinación posterior de resultados. Además, las métricas de exposición deberían expresarse de una forma que permita realizar análisis dosis-respuesta o que pueda transformarse adecuadamente para este propósito, procurando que exista una relación más directa entre la variable utilizada y la exposición relevante para el desenlace reproductivo, por ejemplo, mediante minutos u horas de utilización en una localización corporal pertinente.

En síntesis, los autores consideran que son necesarios estudios prospectivos realizados con mayor rigor metodológico para ampliar la base de evidencia y aumentar la certeza sobre los posibles efectos de la exposición a CEM-RF en los resultados reproductivos femeninos.

La prioridad no reside únicamente en aumentar el número de estudios, sino en mejorar la medición de la exposición, homogeneizar la definición de los desenlaces, ampliar las fuentes y poblaciones estudiadas, incorporar adecuadamente la exposición paterna y garantizar un control y reporte más completos de las variables relevantes.

3.3. Fertilidad masculina: estudios experimentales en mamíferos no humanos y espermatozoides humanos *in vitro*

3.3.1. Referencia completa del artículo

Título: *Effects of radiofrequency electromagnetic field (RF-EMF) exposure on male fertility: A systematic review of experimental studies on non-human mammals and human sperm in vitro.* Cordelli E., Ardoino L., Benassi B., Consales C., Eleuteri P., Marino C., Sciortino M., Villani P., Brinkworth M.H., Chen G., McNamee J.P., Wood A.W., Belackova L., Verbeek J., Pacchierotti F. **Revista y datos bibliográficos:** *Environment International*, 185, 108509 (2024). DOI: 10.1016/j.envint.2024.108509. **Registro de protocolo:** PROSPERO CRD42021227729 para los estudios experimentales en animales y Open Science Framework (OSF), DOI: 10.17605/OSF.IO/7MUS3, para los estudios *in vitro* con espermatozoides humanos.

Título: *Corrigendum to "Effects of radiofrequency electromagnetic field (RF-EMF) exposure on male fertility: A systematic review of experimental studies on non-human mammals and human sperm in vitro" [Environ. Int. 185 (2024) 108509].* Cordelli E., Ardoino L., Benassi B., Consales C., Eleuteri P., Marino C., Sciortino M., Villani P., Brinkworth M.H., Chen G., McNamee J.P., Wood A.W., Belackova L., Verbeek J., Pacchierotti F. **Revista y datos bibliográficos:** *Environment International*, 199, 109449 (2025). DOI: 10.1016/j.envint.2025.109449.

3.3.2. Objetivos y preguntas de investigación

El objetivo de la revisión de Cordelli et al. (2024) fue evaluar de forma sistemática los efectos de la exposición a campos electromagnéticos de radiofrecuencia (CEM-RF) sobre la fertilidad masculina a partir de evidencia experimental. Para ello, los autores integraron dos tipos de estudios complementarios: experimentos *in vivo* en mamíferos no humanos y experimentos *in vitro* con espermatozoides humanos. La revisión se planteó de forma amplia, incluyendo no sólo medidas directas de fertilidad y calidad seminal, sino también diversos desenlaces relacionados con un posible deterioro de la función reproductiva masculina.

En concreto, la revisión se estructuró en torno a dos preguntas PECO (*Population, Exposure, Comparator, Outcome*) diferenciadas:

1. **Estudios en mamíferos no humanos:** ¿cuál es el efecto de la exposición a CEM-RF sobre la fertilidad masculina, en comparación con un grupo de exposición simulada (*sham*) o con un grupo control sometido a la misma temperatura?
2. **Estudios *in vitro* con semen humano:** ¿cuál es el efecto de la exposición a CEM-RF sobre la fertilidad masculina, inferido a partir de estudios con semen humano expuesto *in vitro*, en comparación con un control *sham* o con un control mantenido a la misma temperatura?

Esta doble formulación permite distinguir entre los efectos observados sobre el sistema reproductor de un organismo completo y los detectados directamente en espermatozoides humanos bajo condiciones experimentales controladas. Los desenlaces considerados abarcaron cuatro grandes ámbitos: disminución de la fertilidad; alteraciones de la calidad seminal; toxicidad de los órganos reproductores; y efectos hormonales. En los estudios animales se incluyeron, entre otros, la proporción de machos infértiles o de hembras no gestantes tras el apareamiento, el tamaño de camada, el recuento, vitalidad y morfología espermáticos, las alteraciones del ADN/cromatina espermática, el peso y las alteraciones histológicas e histomorfométricas de testículo y epidídimo, la muerte celular testicular, la producción espermática estimada y los niveles de testosterona. En los estudios *in vitro* con espermatozoides humanos se evaluaron los desenlaces aplicables a este modelo: morfología, vitalidad y alteraciones del ADN/cromatina espermática.

3.3.3. Metodología

La revisión sistemática se realizó de acuerdo con los protocolos previamente publicados por (Pacchierotti et al., 2021) para los estudios experimentales sobre fertilidad masculina. El protocolo correspondiente a los estudios *in vivo* en mamíferos no humanos fue registrado además en PROSPERO (CRD42021227729), mientras que el correspondiente a los estudios *in vitro* con espermatozoides humanos se registró en Open Science Framework (OSF; DOI: 10.17605/OSF.IO/7MUS3). La revisión siguió las directrices metodológicas de la OMS y del *National Toxicology Program/Office of Health Assessment and Translation* (NTP/OHAT). Los autores señalan expresamente algunas desviaciones respecto al protocolo original, que se documentaron y justificaron en la propia revisión.

Búsqueda y fuentes

La búsqueda bibliográfica se realizó en tres bases de datos: PubMed, Scopus y EMF-Portal, esta última especializada en estudios sobre campos electromagnéticos. Las tres fuentes fueron consultadas por última vez el 17 de septiembre de 2022. No se establecieron restricciones por año de publicación ni por idioma.

En PubMed y Scopus se emplearon estrategias construidas a partir de términos en inglés correspondientes a los tres componentes fundamentales de la búsqueda: exposición, población y desenlace. Los términos se combinaron mediante operadores booleanos AND, OR y NOT. La selección de los términos se realizó mediante la revisión de encabezamientos MeSH asociados a artículos relevantes y mediante un proceso iterativo basado en el juicio experto. La exclusión de estudios no experimentales y de estudios en humanos distintos de los experimentos *in vitro* con semen se efectuó manualmente, en lugar de mediante filtros automáticos, para evitar pérdidas derivadas de una posible indexación incorrecta. En EMF-Portal se utilizaron los dominios predefinidos de temática, intervalo de frecuencias y periodo temporal, combinados con términos seleccionados de su glosario. La estrategia de búsqueda había sido sometida a revisión por pares durante el proceso de publicación del protocolo.

Criterios de elegibilidad

Los criterios de elegibilidad se establecieron a priori mediante dos marcos PECO complementarios, uno para experimentos *in vivo* en mamíferos no humanos y otro para experimentos *in vitro* con espermatozoides humanos. Se incluyeron exclusivamente estudios experimentales originales y controlados publicados en revistas sometidas a revisión por pares. Se excluyeron los estudios no experimentales, incluidos los epidemiológicos u observacionales en humanos, así como revisiones, artículos de opinión, actas de congresos y resúmenes de reuniones científicas. También se excluyeron los experimentos en los que se exponían simultáneamente machos y hembras de una pareja reproductora. No se impusieron restricciones por año de publicación ni por idioma.

En los estudios animales, la población elegible estaba constituida por machos de modelos experimentales mamíferos expuestos en cualquier etapa de la vida. Se excluyeron seres humanos y modelos animales no mamíferos. En los estudios *in vitro*, únicamente se admitieron espermatozoides humanos, excluyéndose espermatozoides de otras especies y otros tipos celulares.

La exposición de interés comprendió CEM-RF en el intervalo de 100 kHz a 300 GHz, a cualquier nivel de exposición, y se consideraron asimismo los pulsos electromagnéticos (EMP). Se excluyeron los campos eléctricos o magnéticos estáticos o de frecuencia extremadamente baja, la radiación óptica, los ultrasonidos, la resonancia magnética y las coexposiciones a CEM-RF junto con otros agentes químicos o físicos. También se establecieron criterios específicos para excluir señales cuya energía quedase sustancialmente fuera del intervalo de frecuencias considerado y condiciones en las que no pudiera garantizarse un contraste suficiente de exposición entre el grupo experimental y el comparador.

Los comparadores elegibles fueron controles con exposición simulada (*sham*) o controles mantenidos a igual temperatura. Los controles históricos fueron excluidos. Esta elección resultaba especialmente relevante porque permitía, al menos conceptualmente, distinguir los efectos atribuibles a la exposición a CEM-RF de aquellos derivados únicamente de un incremento térmico.

Los desenlaces animales se agruparon en cuatro categorías. La primera correspondió a disminución de la fertilidad e incluyó proporción de machos infértiles, proporción de hembras no gestantes respecto a las apareadas, tamaño de camada y tasa de fecundación *in vitro*. La segunda comprendió parámetros de calidad seminal: recuento, morfología y vitalidad espermáticos y alteraciones del ADN o de la cromatina. La tercera agrupó indicadores de toxicidad de los órganos reproductores: peso de testículos o epidídimos, histomorfometría e histopatología testicular o epididimaria, muerte celular testicular y producción espermática testicular. Finalmente, se incluyeron los niveles de testosterona como indicador de efectos hormonales. Para los experimentos *in vitro* con espermatozoides humanos se consideraron los desenlaces aplicables a este modelo, fundamentalmente morfología, vitalidad y alteraciones del ADN o cromatina espermática. Se excluyeron los desenlaces que no se consideraron predictivos de deterioro de la fertilidad masculina o que habían sido medidos mediante métodos considerados no válidos.

Una desviación respecto al protocolo afectó a la caracterización de la exposición. Aunque inicialmente se habían establecido criterios más restrictivos, los autores decidieron evaluar también los estudios en los que el nivel de exposición sólo podía inferirse a partir de las condiciones experimentales asumidas y no había sido determinado mediante medición o modelización numérica, debido a la dificultad de establecer un límite inequívoco dentro del continuo de calidad de la información dosimétrica disponible.

Selección y extracción de datos

La selección se realizó en dos etapas. Dos revisores evaluaron independientemente los títulos y resúmenes de los registros identificados. Cuando existía desacuerdo o el resumen no proporcionaba información suficiente para determinar la elegibilidad, el artículo pasaba a la evaluación del texto completo. Posteriormente, dos revisores evaluaron de forma independiente los textos completos y las discrepancias se resolvieron mediante discusión o mediante la intervención de un tercer revisor. Cuando diferentes artículos comunicaban resultados procedentes de un mismo estudio experimental, se consideraron como un único estudio.

La extracción de las características de los estudios y de los datos de los desenlaces fue realizada por dos revisores. La información recopilada incluyó las características del modelo experimental o de las muestras de semen humano, las condiciones de exposición, la dosimetría disponible, los grupos comparadores y los datos necesarios para calcular las medidas de efecto correspondientes.

Medidas de efecto y síntesis cuantitativa

La síntesis cuantitativa se realizó mediante metaanálisis de efectos aleatorios. Los autores efectuaron, por una parte, análisis que comparaban la exposición media con la ausencia de exposición y, por otra, metaanálisis dosis-respuesta que incorporaban los diferentes niveles de exposición disponibles. Dependiendo de la naturaleza del desenlace, los efectos se expresaron mediante medidas apropiadas para variables dicotómicas o continuas, incluidas *odds ratios* (OR), diferencias de medias (MD) y diferencias de medias estandarizadas (SMD). Para las variables dicotómicas se utilizaron OR en lugar de los RR previstos inicialmente en el protocolo. Además, en el análisis de la proporción de hembras no gestantes respecto a las apareadas se aplicó una corrección por agrupamiento intramacho, utilizando un coeficiente de correlación intraclase de 0,2. En la versión corregida de la revisión, cuando un mismo grupo de machos expuestos y su correspondiente grupo *sham* fueron apareados en diferentes momentos con distintos grupos de hembras, los eventos de gestación de las sucesivas rondas de apareamiento se agregaron como un único resultado, evitando así considerar cada apareamiento como un experimento independiente y otorgar un peso excesivo a estos estudios.

Una característica especialmente relevante de esta revisión fue el tratamiento de los experimentos con varios niveles de exposición. Para el análisis principal, los distintos grupos expuestos de un mismo experimento podían combinarse para obtener una exposición media que se comparaba con el correspondiente grupo control. De forma complementaria, los análisis dosis-respuesta utilizaron todos los niveles de exposición disponibles para investigar si la magnitud del efecto variaba sistemáticamente con la exposición. No obstante, los propios autores precisan que determinar la forma exacta de la relación dosis-respuesta o identificar un nivel mínimo de exposición capaz de producir efectos quedaba fuera del objetivo de la revisión.

La heterogeneidad entre estudios se evaluó estadísticamente y se exploraron posibles fuentes mediante análisis de subgrupos. En los estudios experimentales con animales, los subgrupos definidos a priori consideraron fundamentalmente la especie, la SAR y el incremento de temperatura asociado a la exposición.

En los estudios *in vitro* con espermatozoides humanos se había previsto agrupar los resultados según el estado de fertilidad de los donantes y la SAR. Sin embargo, el pequeño número de estudios *in vitro* impidió realizar estos análisis, porque siempre quedaba al menos un subgrupo constituido por menos de tres estudios.

Esta distinción entre SAR e incremento térmico es metodológicamente importante para la revisión, ya que uno de sus objetivos era valorar si la heterogeneidad de los resultados podía explicarse por las condiciones de exposición y, en particular, por exposiciones capaces de producir calentamiento. Los análisis de subgrupos consiguieron explicar parcialmente la heterogeneidad en aproximadamente un tercio de los metaanálisis animales, pero ni estos análisis ni los ajustes dosis-respuesta mostraron una relación consistente entre los niveles de exposición y los efectos observados.

La posibilidad de sesgo de publicación o de notificación se examinó mediante gráficos en embudo (*funnel plots*) y, cuando el número de estudios lo permitía, mediante la prueba de Egger. Esta evaluación no fue posible con igual robustez para todos los desenlaces, especialmente cuando el número de experimentos disponibles era reducido.

Evaluación del riesgo de sesgo (RoB)

El riesgo de sesgo se evaluó siguiendo las directrices OHAT del NTP, adaptadas a las características de los estudios experimentales incluidos. La valoración consideró distintos dominios relacionados con la validez interna, prestando especial atención a la asignación de los animales o muestras a los grupos experimentales, el cegamiento, la caracterización de la exposición, la evaluación de los desenlaces, la integridad de los datos y la notificación selectiva.

A partir de las valoraciones por dominios, los estudios se clasificaron globalmente en tres categorías de preocupación por riesgo de sesgo: “low concern”, “some concern” y “high concern”. Esta clasificación tuvo consecuencias directas sobre la síntesis y, especialmente, sobre la evaluación de la certeza: los estudios clasificados como *high concern* se mostraron y consideraron en los análisis generales, pero fueron excluidos del conjunto utilizado para establecer la certeza de la evidencia mediante GRADE.

La evaluación del riesgo de sesgo resulta particularmente importante en esta revisión por la calidad desigual de la caracterización de la exposición. Los autores prestaron especial atención a si la exposición había sido adecuadamente generada, medida o modelizada y a la fiabilidad de la dosimetría comunicada, incluida la SAR. También fue relevante la evaluación ciega de los desenlaces, aspecto que los propios autores identificaron posteriormente como una de las principales necesidades de mejora de los futuros estudios experimentales.

Certeza de la evidencia (GRADE)

La certeza del cuerpo de evidencia se evaluó mediante el enfoque GRADE, aplicado después de excluir los estudios clasificados como de “high concern” por riesgo de sesgo. La valoración se realizó separadamente para los diferentes desenlaces y tuvo en cuenta los dominios habituales de riesgo de sesgo, inconsistencia, indirectitud, imprecisión y sesgo de publicación. Además, cuando procedía, se consideró la consistencia entre especies como posible factor de incremento de la certeza.

La indirectitud se valoró atendiendo a la relación de cada desenlace experimental con la fertilidad masculina humana. Por este motivo, no todos los indicadores recibieron el mismo peso inferencial: determinados parámetros se consideraron más directamente relacionados con infertilidad masculina que otros. La revisión asignó además a cada desenlace una puntuación de importancia en relación con la infertilidad masculina humana, en una escala de 1 a 10. En los estudios *in vitro* se añadió una fuente inherente de indirectitud, dado que los efectos observados sobre espermatozoides aislados sólo permiten inferir indirectamente posibles efectos sobre la fertilidad de los varones.

Por tanto, a diferencia de las revisiones observacionales precedentes, la certeza no dependió únicamente de la magnitud y precisión de las estimaciones agregadas, sino también de la capacidad de los modelos experimentales y de cada desenlace para informar sobre la fertilidad masculina humana, de la calidad de la dosimetría, de la heterogeneidad entre experimentos y, cuando podía evaluarse, de la posible existencia de sesgo de publicación. Esta estrategia permitió asignar niveles de certeza diferenciados a los distintos desenlaces, en lugar de emitir una valoración única para el conjunto de la evidencia experimental.

3.3.4. Resultados

Flujo PRISMA

El proceso de selección se realizó de forma independiente para los estudios experimentales *in vivo* en mamíferos no humanos y para los estudios *in vitro* con espermatozoides humanos. Finalmente, la revisión incluyó 117 publicaciones sobre estudios experimentales en animales y 10 publicaciones sobre espermatozoides humanos expuestos *in vitro*. Una misma publicación podía contener varios experimentos, niveles de exposición o desenlaces, por lo que los autores distinguieron entre publicaciones y estudios o comparaciones incorporados a las distintas síntesis cuantitativas. Los autores consideraron como un único estudio los resultados procedentes del mismo experimento cuando aparecían distribuidos en varias publicaciones.

El conjunto de evidencia fue particularmente amplio para determinados parámetros seminales y de toxicidad reproductiva. En animales se realizaron síntesis para 14 desenlaces pertenecientes a cuatro categorías: disminución de la fertilidad, calidad seminal, toxicidad de los órganos reproductores y efectos hormonales. En espermatozoides humanos *in vitro* se sintetizaron tres desenlaces de calidad seminal: morfología, vitalidad y alteraciones del ADN/cromatina. Algunos resultados que no podían incorporarse al metaanálisis por haberse evaluado en una sola publicación o mediante métricas no comparables se conservaron en la síntesis narrativa.

Características de los estudios incluidos

Las publicaciones sobre experimentación animal abarcaron un periodo muy amplio, desde 1962 hasta 2022. La gran mayoría utilizaron roedores: 76 publicaciones estudiaron ratas y 38 ratones, mientras que sólo una utilizó hámsteres, una cobayas y una conejos.

La exposición se produjo predominantemente durante la edad adulta, aunque 14 estudios incluyeron exposición total o parcialmente prenatal y 15 exposición total o parcialmente anterior a la pubertad. Las publicaciones sobre espermatozoides humanos *in vitro* fueron publicadas entre 1980 y 2016; casi todas utilizaron muestras procedentes de donantes fértiles, mientras que tres comunicaron resultados en muestras de donantes subfértiles.

Las condiciones experimentales fueron extremadamente heterogéneas. El 97 % de los estudios animales utilizó frecuencias entre 100 MHz y 10 GHz y el 81 % se concentró entre 900 y 2450 MHz, intervalo que incluye frecuencias utilizadas en telefonía móvil y sistemas Wi-Fi. Sólo cinco publicaciones estudiaron frecuencias iguales o superiores a 6000 MHz. La exposición se expresó generalmente mediante SAR, comunicada por los autores o estimada a partir de las características experimentales y de la información dosimétrica disponible. Los valores abarcaron un intervalo extraordinariamente amplio, desde 0,000012 hasta 184 W/kg, y los estudios más antiguos incluían con frecuencia exposiciones muy elevadas dirigidas al estudio de efectos hipotérmicos. La duración de la exposición varió igualmente desde un día hasta un año. Sólo 17 de las 117 publicaciones animales evaluaron más de un nivel de exposición, circunstancia que limitó la información disponible para establecer relaciones dosis-respuesta.

La SAR utilizada correspondía a valores promedio de cuerpo entero incluso en algunos experimentos con exposición localizada, por lo que los autores advierten de que debe interpretarse como un *proxy* del nivel de exposición del tejido diana y no como un indicador directo de la temperatura testicular. No se observó una relación simple entre SAR de cuerpo entero y temperatura de los testículos, ya que ésta podía depender también de la frecuencia, duración, localización de la exposición y temperatura ambiental. Siete publicaciones animales estudiaron específicamente pulsos electromagnéticos (EMP).

En los estudios *in vitro*, ocho de las diez publicaciones utilizaron frecuencias correspondientes a comunicaciones de telefonía móvil. La exposición duró generalmente entre una y unas pocas horas, aunque un estudio alcanzó 16 horas. Sólo una publicación fue diseñada específicamente para explorar una relación dosis-respuesta, lo que restringió considerablemente la posibilidad de investigar gradientes de exposición en este conjunto de evidencia.

Fertilidad en estudios experimentales con animales

La fertilidad se evaluó mediante varios indicadores directos: proporción de machos infértiles, proporción de hembras que no quedaron gestantes después del apareamiento con machos expuestos, tamaño de las camadas engendradas y, en un único estudio, fecundación *in vitro*.

Para la proporción de machos infértiles, los resultados no proporcionaron evidencia de un efecto estadísticamente significativo de la exposición a CEM-RF. El desenlace basado en la proporción de hembras no gestantes tras aparearse con machos expuestos mostró un incremento estadísticamente significativo, con una OR agrupada de 1,68 (IC95 % 1,06–2,65), compatible con una reducción de la capacidad reproductiva masculina.

La certeza de la evidencia para este desenlace se calificó como alta. No obstante, los autores señalan que este resultado está fuertemente condicionado por el estudio de Saunders et al. (1983), realizado con una SAR de 43,4 W/kg; al excluir este estudio, la OR agrupada disminuye a 1,32 y el intervalo de confianza incluye el valor nulo, por lo que consideran cuestionable que exista un efecto perjudicial sobre la tasa de gestación a niveles de exposición relevantes para la población humana. El análisis dosis-respuesta lineal mostró una asociación positiva estadísticamente significativa con la SAR, aunque los análisis globales no permitieron establecer una relación exposición-respuesta consistente ni generalizable.

Para el tamaño de camada, los 15 estudios clasificados con riesgo de sesgo bajo o con algunas preocupaciones proporcionaron una SMD agrupada de 0,04 (IC95 % -0,15 a 0,23), compatible con ausencia de una disminución del número de crías. La SAR media de estos estudios fue muy elevada, 25,83 W/kg, con un intervalo entre 0,31 y 141,4 W/kg. Otros tres estudios clasificados con alta preocupación por riesgo de sesgo proporcionaron una estimación mucho más imprecisa, SMD -0,83 (IC95 % -4,21 a 2,55). Ni los análisis de subgrupos por especie, SAR o incremento de temperatura ni el análisis dosis-respuesta aportaron evidencia consistente de una disminución del tamaño de camada.

La tasa de fecundación *in vitro* sólo fue estudiada en un trabajo, Houston et al. (2019), en ratones expuestos durante 35 días a 2,2 W/kg. No se observaron efectos sobre el porcentaje de ovocitos fecundados ni sobre el desarrollo hasta blastocisto. Por tanto, este desenlace no pudo incorporarse a una síntesis cuantitativa.

Calidad seminal en estudios experimentales con animales

La evidencia fue especialmente abundante para los parámetros seminales. Para el recuento espermático, los resultados individuales presentaban métricas muy diferentes, por lo que fue necesario utilizar la diferencia de medias estandarizada. El metaanálisis de 80 estudios clasificados con riesgo de sesgo bajo o con algunas preocupaciones produjo una SMD de 0,74 (IC95 % 0,51-0,98), indicativa de una disminución estadísticamente significativa de la cantidad de espermatozoides en los animales expuestos. Los siete estudios clasificados con alta preocupación por riesgo de sesgo proporcionaron una SMD de 0,54 (IC95 % -0,06 a 1,15). La SAR media de los primeros fue 12,2 W/kg, con un intervalo de 0,001 a 43,6 W/kg.

Sin embargo, la heterogeneidad del recuento espermático no pudo explicarse mediante los análisis de subgrupos por especie, SAR o incremento de temperatura. Tampoco los modelos dosis-respuesta lineal ni mediante *splines* cúbicos ajustaron satisfactoriamente los resultados. Por ello, aunque el efecto agrupado indica una disminución del recuento espermático, la revisión no demuestra que su magnitud aumente sistemáticamente con la SAR.

Para la morfología espermática, 65 estudios con riesgo bajo o algunas preocupaciones proporcionaron una MD agrupada de -0,94 puntos porcentuales (IC95 % -1,28 a -0,59), indicativa de un incremento estadísticamente significativo de espermatozoides morfológicamente anormales. Los autores destacan, no obstante, que la magnitud absoluta era modesta: el porcentaje medio de espermatozoides anormales en los controles era aproximadamente del 7,7 %, con una variabilidad entre estudios de 0 a 24,6 %. Los cinco estudios clasificados con alta preocupación por riesgo de sesgo mostraron un efecto considerablemente mayor, MD -5,62 (IC95 % -8,75 a -2,50).

En este desenlace, los análisis de subgrupos por especie, SAR e incremento de temperatura sí explicaron parcialmente la heterogeneidad entre estudios. Los modelos dosis-respuesta sugerían asimismo un aumento de las alteraciones morfológicas al incrementarse la exposición, pero su ajuste estadístico fue deficiente, por lo que los autores no consideran demostrada una relación dosis-respuesta fiable.

Para la vitalidad espermática, definida mediante el porcentaje de espermatozoides inmóviles o muertos, 32 estudios con riesgo bajo o algunas preocupaciones proporcionaron una MD agrupada de $-10,83$ puntos porcentuales (IC95 % $-15,20$ a $-6,47$), compatible con una disminución estadísticamente significativa de la vitalidad tras la exposición. La incidencia media basal de espermatozoides inmóviles o muertos en los controles fue aproximadamente del 32 %, aunque con una gran variabilidad entre estudios, del 8 al 53 %. La SAR media fue 1,5 W/kg. Los seis estudios con alta preocupación por riesgo de sesgo mostraron una diferencia todavía mayor, MD $-18,74$ (IC95 % $-29,33$ a $-8,16$).

Los análisis por especie y SAR no explicaron la heterogeneidad observada en vitalidad. Los modelos dosis-respuesta mostraron coeficientes estadísticamente significativos compatibles con un efecto dependiente de la dosis, pero ambos modelos presentaron un ajuste deficiente, de modo que tampoco aquí puede inferirse una relación exposición-respuesta robusta.

Además, el análisis dosis-respuesta lineal mostró una asociación positiva estadísticamente significativa con la SAR, aunque los análisis globales no permitieron establecer una relación exposición-respuesta consistente ni generalizable.

Toxicidad de los órganos reproductores y testosterona

La revisión evaluó también varios indicadores de toxicidad testicular o epididimaria. Para el peso de testículos o epidídimos, los 55 estudios con riesgo bajo o algunas preocupaciones proporcionaron una SMD agrupada de 0,29 (IC95 % 0,10–0,47), compatible con una pequeña disminución del peso en los animales expuestos. Los 13 estudios con alta preocupación por riesgo de sesgo proporcionaron una SMD de 0,42 (IC95 % $-0,03$ a 0,87). Los análisis por especie, SAR o incremento de temperatura no explicaron la heterogeneidad. El modelo lineal sugirió una disminución dependiente de la dosis, pero su pendiente no fue estadísticamente significativa, mientras que el modelo mediante *splines* sugirió un cambio de dirección alrededor de 5 W/kg, sin que ello permitiera establecer un patrón dosis-respuesta consistente.

Para la histomorfometría testicular, 24 estudios con riesgo bajo o algunas preocupaciones mostraron una SMD de 0,90 (IC95 % 0,32–1,49), indicativa de una reducción del diámetro de los túbulos seminíferos. Los 17 estudios con alta preocupación por riesgo de sesgo produjeron una estimación similar, SMD 0,84 (IC95 % 0,27–1,40). La SAR no explicó la heterogeneidad y no pudieron realizarse algunos análisis de subgrupos por especie o temperatura debido al reducido número de estudios en determinadas categorías.

Además, el análisis dosis-respuesta lineal mostró una asociación positiva estadísticamente significativa con la SAR, aunque los análisis globales no permitieron establecer una relación exposición-respuesta consistente ni generalizable.

Los estudios sobre testosterona mostraron resultados heterogéneos. En los estudios con riesgo de sesgo bajo o con algunas preocupaciones, el metaanálisis mostró una SMD de 0,92 (IC95 % 0,49–1,35), en la dirección correspondiente a una reducción de los niveles de testosterona en los animales expuestos.

Sin embargo, los análisis exposición-respuesta no mostraron un patrón monotónico consistente: el ajuste mediante *splines* sugirió una relación compleja, con cambios de dirección a distintos niveles de SAR. Por ello, los autores no identificaron una relación dosis-respuesta clara y la certeza de la evidencia se calificó como muy baja.

Estudios con espermatozoides humanos expuestos in vitro

La evidencia *in vitro* fue mucho más reducida que la animal. Se evaluaron fundamentalmente morfología, vitalidad y alteraciones del ADN/cromatina. Para morfología espermática, sólo un estudio no clasificado como de alta preocupación por riesgo de sesgo aportó información, por lo que la evidencia fue insuficiente para obtener una estimación robusta.

La MD agrupada fue de $-1,37$ puntos porcentuales (IC95 % $-2,46$ a $-0,28$). La heterogeneidad fue apreciable ($I^2 = 65$ %) y no pudo explicarse mediante análisis de subgrupos. Además, la interpretación está necesariamente limitada por la indirectitud inherente a extrapolar los efectos observados sobre espermatozoides aislados *in vitro* a la fertilidad de los varones.

En cambio, para las alteraciones del ADN/cromatina, la SMD agrupada fue de $-0,17$ (IC95 % $-0,48$ a $0,13$), por lo que el intervalo incluyó el valor nulo y fue compatible con ausencia de efecto. También aquí existió heterogeneidad ($I^2 = 53$ %) y una importante indirectitud. Por tanto, los estudios *in vitro* proporcionan indicios de una pequeña reducción de la vitalidad, pero no evidencia de alteraciones del ADN/cromatina en la estimación agrupada. La certeza de ambos resultados fue muy baja.

El pequeño número de trabajos impidió realizar los análisis de subgrupos previstos por estado de fertilidad de los donantes y SAR, ya que siempre quedaba al menos un subgrupo constituido por menos de tres estudios. Esta limitación es importante porque impide determinar si las diferencias entre experimentos pueden relacionarse con la condición de los donantes o con el nivel de exposición.

Exposición a pulsos electromagnéticos (EMP) y análisis de exposición-respuesta

Los estudios sobre EMP constituyeron un subconjunto pequeño y se analizaron separadamente. En general, no mostraron efectos consistentes sobre los desenlaces evaluados. Por ejemplo, para la proporción de hembras no gestantes, cinco estudios con riesgo bajo o algunas preocupaciones proporcionaron una OR de $0,89$ (IC95 % $0,38-2,10$), y para tamaño de camada una SMD de $-0,38$ (IC95 % $-0,91$ a $0,15$). Para recuento espermático, el resultado fue asimismo compatible con ausencia de efecto, SMD $0,23$ (IC95 % $-0,09$ a $0,56$). La certeza de estos resultados osciló entre baja y muy baja.

Considerando conjuntamente los estudios animales convencionales, los análisis de subgrupos por especie, SAR e incremento térmico permitieron explicar parcialmente la heterogeneidad en aproximadamente un tercio de los metaanálisis. Sin embargo, los análisis de subgrupos y los ajustes dosis-respuesta no mostraron una relación consistente entre los niveles de exposición y los efectos observados.

Este resultado es fundamental para interpretar las asociaciones estadísticamente significativas encontradas en diversos desenlaces: la existencia de una diferencia agregada entre expuestos y controles no se acompañó de un gradiente exposición-respuesta reproducible.

Además, los niveles de exposición empleados en muchos experimentos fueron superiores a los característicos de la exposición humana habitual e incluso a los límites establecidos por las guías internacionales. Por ello, los propios autores advierten de que sus resultados no permiten establecer un nivel mínimo de exposición capaz de producir efectos ni utilizar directamente esta evidencia para confirmar o reconsiderar los límites actuales de exposición humana.

Evaluación del riesgo de sesgo

El riesgo de sesgo constituye una de las principales limitaciones del cuerpo de evidencia. Sólo una pequeña proporción de los estudios fue clasificada como de "low concern", mientras que la mayoría presentó alguna preocupación y una fracción fue clasificada como de "high concern". Los problemas se concentraron especialmente en la caracterización de la exposición y en la evaluación de los desenlaces, aspectos críticos en estudios experimentales de bioelectromagnetismo.

Las limitaciones en la exposición incluyeron información dosimétrica incompleta, caracterización insuficiente de las condiciones experimentales y, en algunos casos, estimaciones de SAR derivadas indirectamente de la configuración experimental. También fueron frecuentes las deficiencias en el cegamiento de quienes evaluaban los desenlaces. Estas limitaciones son especialmente relevantes porque muchos de los resultados, morfología espermática, histología, histomorfometría o muerte celular, requieren procedimientos de evaluación susceptibles de sesgo cuando el examinador conoce el grupo experimental.

La clasificación global del riesgo de sesgo tuvo consecuencias directas para la interpretación. Los estudios de alta preocupación se representaron separadamente en los *forest plots* y pudieron contribuir a las síntesis generales, pero se excluyeron del cuerpo de evidencia utilizado para establecer la certeza mediante GRADE. Esta separación resulta importante porque, para algunos desenlaces, las estimaciones procedentes de los estudios con alta preocupación por riesgo de sesgo mostraban efectos considerablemente mayores que las obtenidas en los estudios metodológicamente más fiables.

La evaluación del sesgo de publicación añadió otra fuente de preocupación. Los *funnel plots* y las pruebas de Egger indicaron un posible sesgo de publicación para numerosos desenlaces. Las principales excepciones fueron la tasa de gestaciones, el tamaño de camada y la producción espermática testicular en animales, así como las alteraciones del ADN/cromatina en espermatozoides humanos *in vitro*. En algunos desenlaces, sin embargo, el número de estudios fue demasiado pequeño para que la prueba de Egger pudiera aplicarse con fiabilidad.

Certeza de la evidencia (GRADE)

La evaluación GRADE se realizó excluyendo los estudios clasificados con alta preocupación por riesgo de sesgo y produjo un panorama muy desigual según el desenlace. La mayoría de las asociaciones observadas recibió una certeza muy baja, debido principalmente al riesgo de sesgo, inconsistencia, indirectitud, imprecisión y posible sesgo de publicación. Sin embargo, hubo tres resultados que alcanzaron niveles de certeza superiores y que, por ello, tienen especial relevancia para la interpretación de la revisión.

La evidencia de una reducción de la tasa de gestación, evaluada mediante la proporción de hembras que no quedaron gestantes tras aparearse con machos expuestos, recibió una certeza alta. También se asignó certeza moderada a la evidencia de ausencia o escaso efecto sobre el tamaño de las camadas. Por su parte, la disminución del recuento espermático recibió una certeza baja. Estos son los resultados con mayor confianza de toda la revisión.

En contraste, los efectos observados sobre morfología y vitalidad espermáticas, alteraciones del ADN/cromatina, peso de testículos o epidídimos, histomorfometría e histopatología testicular, muerte celular, producción espermática y testosterona fueron considerados de certeza muy baja. En varios de estos desenlaces existían estimaciones estadísticamente significativas orientadas hacia un efecto adverso, pero la elevada heterogeneidad, las limitaciones metodológicas, la indirectitud respecto a infertilidad masculina humana y el posible sesgo de publicación impidieron otorgarles una confianza mayor. Por ejemplo, para histomorfometría testicular la heterogeneidad alcanzó $I^2 = 84 \%$, y para algunos indicadores se consideró además débil su capacidad predictiva respecto a infertilidad humana.

En los estudios con espermatozoides humanos *in vitro*, tanto la pequeña reducción de la vitalidad como la ausencia de efecto sobre alteraciones del ADN/cromatina recibieron certeza muy baja. Además de las limitaciones metodológicas e inconsistencia, GRADE penalizó expresamente la indirectitud derivada de utilizar espermatozoides aislados para inferir efectos sobre la fertilidad de los hombres; para las alteraciones del ADN/cromatina se añadió una degradación adicional por la relación relativamente débil de este desenlace con la infertilidad masculina.

Finalmente, la evidencia procedente de los experimentos con EMP fue calificada entre baja y muy baja certeza y, en términos generales, no mostró efectos sobre los desenlaces estudiados.

En conjunto, por tanto, los resultados de Cordelli et al. no pueden resumirse simplemente como evidencia experimental de que los CEM-RF deterioren la fertilidad masculina. Los metaanálisis muestran estimaciones orientadas hacia efectos adversos en numerosos desenlaces, pero la certeza es muy baja para la mayoría de ellos y no aparece una relación dosis-respuesta consistente. Los resultados con mayor respaldo son más específicos: certeza alta de una reducción de la tasa de gestación tras el apareamiento con machos expuestos, aunque este resultado está fuertemente condicionado por un estudio realizado con una SAR extremadamente elevada; certeza moderada de ausencia o escaso efecto sobre el tamaño de camada y certeza baja de una reducción del recuento espermático. A ello se añade que gran parte de la experimentación animal utilizó exposiciones superiores a las habituales en humanos. Estas consideraciones son esenciales para evitar trasladar directamente los efectos experimentales observados a las condiciones reales de exposición humana.

3.3.5. Fortalezas y debilidades de la revisión

Fortalezas

Una de las principales fortalezas de la revisión de Cordelli et al. (2024) es su amplitud y carácter sistemático. A diferencia de revisiones anteriores sobre fertilidad masculina y CEM-RF, que habían restringido el análisis a determinados parámetros seminales, condiciones de exposición o intervalos de SAR, los autores plantearon una evaluación integral de la evidencia experimental disponible. Para ello incluyeron tanto estudios *in vivo* en mamíferos no humanos como estudios *in vitro* con espermatozoides humanos y consideraron múltiples desenlaces relacionados con la función reproductiva masculina: fertilidad propiamente dicha, calidad seminal, toxicidad de los órganos reproductores y efectos hormonales. Esta amplitud permite distinguir entre desenlaces directamente relacionados con la capacidad reproductiva y marcadores intermedios con diferente relevancia predictiva para la infertilidad masculina.

La revisión se sustenta, además, en un protocolo previamente publicado (Pacchierotti et al., 2021) y en una metodología alineada con las directrices de la OMS y el *National Toxicology Program/Office of Health Assessment and Translation* (NTP/OHAT). Los criterios PECO, de elegibilidad y de validez de los métodos analíticos fueron establecidos previamente. La búsqueda se realizó sin restricciones de fecha o idioma en PubMed, Scopus y EMF-Portal, esta última específicamente dedicada a estudios sobre campos electromagnéticos. La estrategia de búsqueda había sido sometida a revisión por pares durante la publicación del protocolo. La selección por título/resumen y posteriormente por texto completo fue realizada de forma independiente por dos revisores, con resolución de discrepancias mediante discusión o participación de un tercero.

Otra fortaleza relevante es el tratamiento explícito de la calidad metodológica de los estudios primarios. La revisión no se limita a agregar los resultados experimentales, sino que incorpora una evaluación formal del riesgo de sesgo y diferencia en los *forest plots* los estudios clasificados con alta preocupación de aquellos con riesgo bajo o algunas preocupaciones. Más importante aún, los estudios con alta preocupación por riesgo de sesgo fueron excluidos del cuerpo de evidencia utilizado para determinar la certeza mediante GRADE. De este modo, los efectos observados en experimentos metodológicamente más débiles no reciben el mismo peso en la interpretación final que los procedentes de estudios considerados más fiables.

También resulta especialmente valioso el esfuerzo por investigar la heterogeneidad y las posibles relaciones exposición-respuesta. Los autores realizaron, cuando los datos lo permitieron, análisis de subgrupos según especie, SAR e incremento de temperatura, además de modelos dosis-respuesta lineales y mediante *splines*. Esta estrategia es particularmente importante en la investigación experimental sobre CEM-RF, porque permite explorar si los efectos observados aumentan con la exposición y si pueden estar relacionados con el calentamiento producido a niveles elevados. El hecho de que estos análisis no identificaran un gradiente dosis-respuesta consistente constituye en sí mismo información relevante para interpretar causalmente los resultados.

Finalmente, la revisión realiza una distinción explícita entre significación estadística, relevancia biológica y certeza de la evidencia. Los autores asignaron a los distintos desenlaces una importancia diferente según su capacidad para predecir infertilidad masculina y aplicaron GRADE teniendo en cuenta riesgo de sesgo, inconsistencia, indirectitud, imprecisión y sesgo de publicación. Esta aproximación evita interpretar automáticamente como evidencia sólida de daño cualquier diferencia estadísticamente significativa encontrada en los metaanálisis y permite situar adecuadamente resultados como las alteraciones histológicas o celulares frente a desenlaces más directamente relacionados con la capacidad reproductiva.

Limitaciones

La principal limitación de la revisión no reside tanto en el procedimiento de revisión como en las características del cuerpo de evidencia disponible. Los estudios incluidos fueron extraordinariamente heterogéneos en especie, frecuencia, modulación, intensidad, duración y modalidad de exposición, etapa vital en la que se produjo ésta y procedimientos empleados para evaluar los desenlaces. Aunque esta diversidad permite cubrir un amplio espacio experimental, dificulta considerablemente la interpretación de los efectos agrupados y explica parte de la elevada heterogeneidad observada en numerosos metaanálisis.

Particularmente importante es la heterogeneidad dosimétrica. Aunque la SAR permitió disponer de una métrica común para muchos estudios, su caracterización fue frecuentemente incompleta y una proporción importante de publicaciones presentó problemas en este dominio. De hecho, algunos trabajos tuvieron que ser excluidos de la revisión por una caracterización insuficiente de la exposición, mientras que una parte considerable de los incluidos recibió una clasificación de «algunas preocupaciones» o «alta preocupación» por razones semejantes. Además, la SAR promedio de cuerpo entero se utilizó en ocasiones como aproximación a la exposición del órgano diana, aunque no necesariamente representa adecuadamente la dosis o el incremento térmico recibido por los testículos.

Una limitación crucial para extrapolar los resultados a humanos es que gran parte de la experimentación animal utilizó niveles de exposición considerablemente superiores a los habituales en la población. Los autores señalan que, dependiendo del grupo de desenlaces, aproximadamente el 75–80 % de los estudios ensayaron niveles superiores a 0,4 W/kg, valor que toman como referencia de la restricción básica de ICNIRP para trabajadores, y alrededor del 32–53 % estudiaron niveles superiores a 4 W/kg, valor utilizado por ICNIRP como umbral de SAR de cuerpo entero asociado a efectos adversos sobre la salud y a partir del cual se derivan las restricciones básicas mediante factores de reducción. En consecuencia, permanece incierto hasta qué punto las estimaciones obtenidas mediante los metaanálisis pueden extrapolarse a los niveles de exposición característicos de la población humana. La generalización a los extremos del intervalo de frecuencias es igualmente limitada, puesto que se recuperaron muy pocos estudios por debajo de 100 MHz o por encima de 10 GHz.

La capacidad para establecer relaciones dosis-respuesta constituye otra limitación importante. Sólo una minoría de los estudios animales evaluó varios niveles de exposición dentro del mismo experimento y únicamente un estudio *in vitro* fue diseñado específicamente para investigar una relación exposición-respuesta. En consecuencia, buena parte de los análisis dosis-respuesta se construyó comparando niveles de exposición entre experimentos realizados en condiciones muy diferentes.

Aunque los autores aplicaron modelos lineales y no lineales, su ajuste fue frecuentemente deficiente y no apareció un patrón exposición-respuesta consistente. Los propios autores señalan expresamente que determinar la forma de la relación dosis-respuesta o identificar un nivel mínimo efectivo quedaba fuera del alcance de la revisión; por ello, los resultados no permiten confirmar ni reconsiderar los límites vigentes de exposición humana.

También existen limitaciones metodológicas recurrentes en los experimentos primarios. Los autores destacan la ausencia de cegamiento durante la realización de los experimentos o la evaluación de los desenlaces, un número insuficiente de preparaciones citológicas o histológicas analizadas, el empleo de procedimientos de evaluación manual potencialmente susceptibles de subjetividad y una descripción incompleta de los métodos experimentales y de los resultados. Estas deficiencias son particularmente relevantes para desenlaces como morfología espermática, histología o histomorfometría, cuya evaluación puede estar influida por el conocimiento del grupo experimental. La revisión identifica además indicios de posible sesgo de publicación para diversos desenlaces.

Debe considerarse asimismo la indirectitud inherente a buena parte de la evidencia. Incluso los estudios animales mejor realizados proporcionan evidencia indirecta respecto a la fertilidad masculina humana, y algunos de los desenlaces experimentales, como histomorfometría, muerte celular o determinadas alteraciones testiculares, tienen una capacidad predictiva limitada respecto a infertilidad. Esta circunstancia explica que varios efectos estadísticamente significativos recibieran finalmente una certeza muy baja. La evidencia *in vitro* presenta una indirectitud todavía mayor: la exposición de espermatozoides aislados no reproduce las condiciones fisiológicas de exposición de los espermatozoides dentro del aparato reproductor masculino ni permite evaluar la fertilidad como fenómeno integrado.

Finalmente, el propio artículo reconoce las limitaciones específicas del modelo *in vitro* con espermatozoides humanos. El reducido número de estudios, la heterogeneidad experimental y la dificultad para extrapolar sus resultados a la fertilidad masculina *in vivo* disminuyen considerablemente su utilidad. La valoración de los autores es suficientemente crítica como para que, en sus recomendaciones de investigación, no aconsejen continuar realizando estudios de este tipo, y propongan en su lugar desarrollar mejores marcadores predictivos de infertilidad en modelos experimentales animales y aprovechar el análisis de calidad seminal en estudios de biomonitorización de poblaciones humanas expuestas a CEM-RF.

En conjunto, Cordelli et al. (2024) constituye una revisión sistemática amplia y metodológicamente estructurada que mejora sustancialmente evaluaciones anteriores al integrar riesgo de sesgo, metaanálisis, análisis de heterogeneidad y exposición-respuesta y valoración GRADE. Sin embargo, la capacidad de la revisión para extraer conclusiones aplicables a la exposición humana está condicionada por las limitaciones de la evidencia primaria: heterogeneidad experimental y dosimétrica, riesgo de sesgo, escasez de experimentos con múltiples niveles de exposición, posibles sesgos de publicación y, de manera especialmente relevante, utilización frecuente de niveles de exposición superiores a los característicos de la población humana. Estas limitaciones explican que, pese a observarse efectos estadísticamente significativos en varios desenlaces, la certeza de la evidencia sea baja o muy baja para la mayoría de ellos y que no pueda establecerse un gradiente exposición-respuesta generalizable.

3.3.6. Conclusiones

En conjunto, la revisión de Cordelli et al. (2024) muestra que la exposición experimental a CEM-RF puede asociarse con alteraciones de determinados parámetros relacionados con la reproducción masculina en mamíferos no humanos, pero la interpretación del conjunto de evidencia requiere distinguir cuidadosamente entre los distintos desenlaces, su relevancia para la fertilidad y la certeza asignada a cada resultado. Los autores subrayan que la evidencia experimental disponible es extensa, pero muy heterogénea, y que para la mayoría de los desenlaces la certeza es baja o muy baja.

Entre los indicadores más directamente relacionados con la capacidad reproductiva, los machos expuestos a CEM-RF presentaron, en el análisis agrupado, una menor tasa de gestación de las hembras con las que fueron apareados, resultado para el que la revisión atribuye una certeza alta. No obstante, los autores señalan que este resultado está fuertemente condicionado por un estudio realizado con una SAR de 43,4 W/kg y que, al excluirlo, la estimación agrupada deja de ser estadísticamente significativa, por lo que consideran cuestionable que exista un efecto perjudicial sobre la tasa de gestación a niveles de exposición relevantes para la población humana. En cambio, no se observó una reducción del tamaño de las camadas, con certeza moderada.

Esta diferencia entre ambos desenlaces es relevante: los resultados indican un efecto sobre la probabilidad de conseguir una gestación en el conjunto de estudios analizados, pero no indican que, una vez establecida ésta, la exposición paterna reduzca de manera apreciable el número de crías. Otros indicadores directos de fertilidad fueron evaluados por un número insuficiente de estudios para alcanzar conclusiones firmes.

En relación con la calidad seminal, el resultado más consistente fue la disminución del recuento espermático, para la que los autores calificaron la certeza como baja. Aunque el recuento espermático no constituye por sí mismo una medida funcional de fertilidad, se trata de un parámetro estandarizado y utilizado rutinariamente en andrología clínica, por lo que los autores consideran que este resultado no debería ser ignorado.

También se observaron efectos adversos estadísticamente significativos sobre otros parámetros, como morfología y vitalidad espermáticas y diversos indicadores de toxicidad de los órganos reproductores, pero la certeza de estas asociaciones fue muy baja. En consecuencia, su significación estadística no debe confundirse con evidencia sólida de un efecto reproductivo.

Un aspecto fundamental para interpretar estos resultados es que los análisis realizados no permitieron establecer una relación exposición-respuesta consistente. Aunque determinados modelos mostraron asociaciones con la SAR para algunos desenlaces, globalmente no apareció un patrón reproducible según el cual el efecto aumentase progresivamente al incrementarse la exposición. La revisión tampoco estaba diseñada para determinar la forma precisa de una eventual curva dosis-respuesta ni para identificar un nivel mínimo de exposición capaz de producir efectos. Por esta razón, los autores indican expresamente que sus resultados no permiten proponer la confirmación ni la reconsideración de los límites actuales de exposición humana.

Esta cautela resulta particularmente importante porque una gran proporción de los experimentos animales utilizó niveles de exposición elevados en comparación con los relevantes para la población humana: dependiendo del grupo de desenlaces, el 75–80 % de los estudios empleó niveles superiores a 0,4 W/kg y el 32–53 % niveles superiores a 4 W/kg, valores utilizados por ICNIRP como umbral de SAR de cuerpo entero asociado a efectos adversos sobre la salud y empleado como punto de partida para derivar las restricciones básicas mediante factores de reducción, ocupacional y público general respectivamente.

Por ello, los propios autores consideran desconocido hasta qué punto las conclusiones de los metaanálisis pueden extrapolarse a los niveles habituales de exposición humana. La extrapolación es igualmente incierta para frecuencias inferiores a 100 MHz o superiores a 10 GHz, debido al reducido número de estudios disponibles en esos intervalos.

La evidencia obtenida mediante exposición *in vitro* de espermatozoides humanos es todavía más limitada. Los resultados sugieren una pequeña disminución de la vitalidad espermática, mientras que la estimación agrupada para alteraciones del ADN/cromatina fue compatible con ausencia de efecto; en ambos casos, la certeza fue muy baja. Además, este modelo presenta una importante indirectitud respecto a la fertilidad masculina, porque la exposición de espermatozoides aislados no reproduce las condiciones fisiológicas existentes en el aparato reproductor. Los propios autores consideran suficientemente importantes estas limitaciones como para no recomendar la realización de nuevos estudios basados en este enfoque.

Por tanto, la conclusión de Cordelli et al. debe interpretarse de forma matizada. La revisión identifica efectos adversos sobre algunos indicadores de fertilidad masculina, particularmente una disminución de la tasa de gestación y del recuento espermático, aunque con niveles de certeza diferentes y sin demostrar un deterioro generalizado y consistente de todos los desenlaces reproductivos ni establecer un gradiente exposición-respuesta robusto. En particular, pese a la certeza alta asignada a la reducción de la tasa de gestación en el análisis agrupado, los autores consideran cuestionable que este efecto se produzca a niveles de exposición relevantes para la población humana, debido a la influencia de un estudio realizado con una SAR extremadamente elevada. Además, la elevada heterogeneidad, el riesgo de sesgo, el posible sesgo de publicación, la certeza baja o muy baja para la mayoría de los desenlaces y, sobre todo, las diferencias entre las condiciones experimentales y los niveles habituales de exposición humana impiden trasladar directamente estos resultados a una estimación del riesgo para las personas.

Implicaciones para investigación futura

Cordelli et al. formulan recomendaciones relativamente concretas para mejorar la investigación experimental futura. En primer lugar, consideran prioritario aumentar la calidad metodológica de los estudios. Recomiendan aplicar sistemáticamente cegamiento tanto durante la realización de los experimentos como durante la evaluación de los desenlaces, utilizar un número suficiente de preparaciones citológicas o histológicas, emplear métodos automatizados de análisis siempre que sea posible y mejorar la estandarización y exhaustividad con que se describen los procedimientos técnicos y los resultados. Estas medidas están directamente relacionadas con las deficiencias que determinaron la exclusión de numerosos trabajos o su clasificación con preocupaciones por riesgo de sesgo.

Una segunda prioridad es mejorar sustancialmente la caracterización de la exposición. Muchos estudios tuvieron que ser excluidos precisamente porque ésta era insuficiente, y problemas semejantes explicaron la clasificación de una proporción importante de los trabajos incluidos como de «algunas preocupaciones» o «alta preocupación» por riesgo de sesgo. Los autores recomiendan que los futuros diseños experimentales incorporen desde su planificación los requisitos necesarios para una dosimetría suficientemente precisa y reproducible.

Especialmente importante es la recomendación de abandonar diseños basados en un único nivel de exposición. Los futuros experimentos deberían estudiar varios niveles dentro de las mismas condiciones experimentales, abarcando desde exposiciones bajas comparables a las humanas hasta niveles superiores en los que puedan esperarse efectos hipotérmicos leves.

Esta estrategia permitiría estudiar de manera mucho más fiable la existencia y la forma de posibles relaciones exposición-respuesta, evitando tener que inferirlas mediante comparaciones entre experimentos realizados en condiciones heterogéneas. Los autores recomiendan, además, monitorizar la temperatura del tejido diana, un aspecto esencial para distinguir posibles efectos específicos de los CEM-RF de los derivados del calentamiento.

Como prioridad final, los autores proponen desarrollar una base experimental científicamente sólida sobre los efectos de los CEM-RF en los marcadores sustitutivos con mayor capacidad predictiva de infertilidad masculina en roedores. Sólo a partir de una mejor identificación y validación de estos marcadores consideran que podría plantearse posteriormente estudiar de manera directa el impacto de los CEM-RF sobre el rendimiento reproductivo masculino. Esta propuesta busca concentrar los recursos en desenlaces con verdadera relevancia funcional, en lugar de acumular resultados sobre alteraciones celulares, histológicas o seminales cuya relación con la fertilidad puede ser incierta.

En cuanto a la investigación con material humano, Cordelli et al. hacen una recomendación particularmente clara: debido a las limitaciones del modelo de exposición *in vitro* de espermatozoides humanos, no recomiendan realizar más estudios de este tipo. Como alternativa, proponen aprovechar el análisis de la calidad seminal en estudios de biomonitorización de poblaciones humanas expuestas a CEM-RF. De este modo, la investigación futura podría aproximarse mejor a condiciones reales de exposición y complementar la evidencia experimental animal con información directamente obtenida en seres humanos.

En síntesis, la prioridad no sería simplemente aumentar el número de experimentos, sino producir estudios mejor diseñados, con dosimetría rigurosa, cegamiento, evaluación objetiva de los desenlaces, varios niveles de exposición bajo condiciones experimentales comparables y control de la temperatura tisular. Sólo mediante este tipo de diseños podrá determinarse si las señales observadas en algunos parámetros reproductivos responden a efectos reproducibles de los CEM-RF, establecerse una relación exposición-respuesta fiable y, sobre todo, aclararse su relevancia para los niveles de exposición que realmente experimenta la población humana.

3.4. Embarazo y resultados al nacimiento: estudios experimentales en mamíferos no humanos

3.4.1. Referencia completa del artículo

Título: *Effects of Radiofrequency Electromagnetic Field (RF-EMF) exposure on pregnancy and birth outcomes: A systematic review of experimental studies on non-human mammals.* Cordelli E., Ardoino L., Benassi B., Consales C., Eleuteri P., Marino C., Sciortino M., Villani P., Brinkworth M.H., Chen G., McNamee J.P., Wood A.W., Belackova L., Verbeek J., Pacchierotti F. **Revista y datos bibliográficos:** *Environment International*, 180, 108178 (2023). DOI: 10.1016/j.envint.2023.108178. **Registro de protocolo:** PROSPERO CRD42021227746.

3.4.2. Objetivos y preguntas de investigación

El objetivo principal de Cordelli et al. (2023) fue evaluar sistemáticamente la evidencia experimental disponible sobre los posibles efectos adversos de la exposición a campos electromagnéticos de radiofrecuencia (CEM-RF) durante la gestación sobre el embarazo, el desarrollo embrionario y fetal y la salud de la descendencia, utilizando estudios realizados en mamíferos no humanos. La revisión se desarrolló en el marco del proyecto de la OMS para la evaluación sistemática de los posibles efectos sobre la salud de la exposición a CEM-RF y pretendía superar las limitaciones de evaluaciones anteriores mediante una revisión sistemática conforme a las directrices de la OMS y del *National Toxicology Program/Office of Health Assessment and Translation* (NTP/OHAT).

La pregunta de investigación se estructuró mediante un esquema PECO (*Population, Exposure, Comparator, Outcome*). La población (P) estuvo constituida por mamíferos experimentales expuestos exclusivamente *in utero*. La exposición (E) comprendió CEM-RF en el intervalo de frecuencias de 100 kHz a 300 GHz. Como comparadores (C) se consideraron animales sometidos a exposición simulada (*sham-exposed*) o animales en los que se reprodujo mediante calentamiento directo el mismo incremento de temperatura inducido por la exposición a CEM-RF. Los desenlaces (O) se organizaron en tres grandes categorías: reducción de la fecundidad, efectos adversos sobre la salud de la descendencia al nacimiento y efectos adversos de aparición posterior en la descendencia; cada categoría comprendía varios desenlaces específicos no redundantes.

En concreto, la reducción de la fecundidad se evaluó mediante parámetros relacionados con el tamaño de la camada y las pérdidas embrionarias o fetales. Los efectos adversos presentes al nacimiento comprendieron disminuciones del peso o de la longitud fetal, malformaciones congénitas y alteraciones de la proporción de sexos. Finalmente, los efectos diferidos sobre la salud de la descendencia incluyeron alteraciones neurocognitivas, efectos sobre la fertilidad de las hembras descendientes y cáncer de aparición temprana.

Por tanto, la cuestión central de la revisión fue determinar si la exposición gestacional a CEM-RF en mamíferos experimentales produce efectos adversos sobre la fecundidad, el desarrollo prenatal, la salud al nacimiento o la salud posterior de la descendencia, y, cuando los datos disponibles lo permitieran, caracterizar cuantitativamente esos efectos y su relación con el nivel de exposición. Esta última dimensión resulta especialmente relevante porque la revisión no se planteó únicamente establecer diferencias entre animales expuestos y controles, sino también examinar las relaciones exposición-respuesta mediante metaanálisis dosis-respuesta.

3.4.3. Metodología

La revisión sistemática se realizó de acuerdo con el protocolo previamente publicado por Pacchierotti et al. (2021) para las revisiones sistemáticas de estudios experimentales sobre fertilidad masculina y embarazo y resultados al nacimiento. El protocolo correspondiente a los estudios *in vivo* en mamíferos no humanos fue registrado además en PROSPERO (CRD42021227746) para los estudios *in vivo* en mamíferos no humanos. La revisión siguió las directrices metodológicas de la OMS y del *National Toxicology Program/Office of Health Assessment and Translation* (NTP/OHAT). Los autores señalan expresamente algunas desviaciones respecto al protocolo original, que se documentaron y justificaron en la propia revisión.

Búsqueda y fuentes

La búsqueda bibliográfica se realizó en tres bases de datos: PubMed, Scopus y EMF-Portal, esta última especializada en estudios sobre campos electromagnéticos. PubMed y Scopus fueron consultadas por última vez el 8 de septiembre de 2022 y EMF-Portal el 17 de septiembre de 2022. No se establecieron restricciones por año de publicación ni por idioma.

En PubMed y Scopus se utilizaron estrategias construidas a partir de términos en inglés correspondientes a los tres componentes fundamentales de la búsqueda: exposición, población y desenlace. Estos términos se combinaron mediante operadores booleanos AND, OR y NOT. Su selección se realizó mediante la revisión de encabezamientos MeSH asociados a publicaciones relevantes y mediante un proceso iterativo basado en el juicio experto. La exclusión de estudios no experimentales y de estudios en humanos se efectuó manualmente, en lugar de mediante filtros automáticos, para reducir el riesgo de excluir trabajos incorrectamente indexados. En EMF-Portal se utilizaron los dominios predefinidos de temática, intervalo de frecuencias y periodo temporal, combinados con términos seleccionados de su glosario. La estrategia de búsqueda había sido sometida a revisión por pares durante el proceso de publicación del protocolo.

Criterios de elegibilidad

Los criterios de elegibilidad se establecieron *a priori* mediante el marco PECO. Se incluyeron exclusivamente estudios experimentales originales y controlados realizados en mamíferos no humanos expuestos *in utero* y publicados en revistas sometidas a revisión por pares.

Se excluyeron los estudios no experimentales, incluidos los epidemiológicos u observacionales en humanos, así como revisiones, artículos de opinión, actas de congresos y resúmenes de reuniones científicas. No se impusieron restricciones por año de publicación ni por idioma.

La población elegible estuvo constituida por mamíferos experimentales expuestos exclusivamente *in utero*. Por tanto, la exposición debía producirse durante el desarrollo prenatal; quedaron fuera del objeto específico de esta revisión los experimentos limitados a exposiciones preconcepcionales o exclusivamente posnatales. La exposición de interés comprendió CEM-RF en el intervalo de frecuencias de 100 kHz a 300 GHz. Los comparadores elegibles fueron animales sometidos a exposición simulada (*sham*) o animales expuestos mediante calentamiento directo al mismo incremento de temperatura inducido por los CEM-RF.

Los desenlaces se organizaron en tres categorías principales: reducción de la fecundidad, efectos adversos sobre la salud de la descendencia al nacimiento y efectos adversos diferidos sobre la salud de la descendencia. Cada una de estas categorías comprendió varios desenlaces específicos no redundantes. Entre los indicadores de reducción de la fecundidad se consideraron las pérdidas preimplantacionales, el tamaño de la camada y las pérdidas fetales o resorciones. Los efectos al nacimiento incluyeron parámetros de crecimiento fetal, como peso y longitud, malformaciones fetales y proporción de sexos. Los efectos diferidos comprendieron indicadores de alteración del sistema nervioso central y del desarrollo neuroconductual, funciones de aprendizaje y memoria, actividad motora, funciones motoras y sensoriales y fertilidad de las hembras de la descendencia. La carcinogenicidad en la descendencia se había contemplado dentro de los posibles efectos diferidos, aunque finalmente ningún estudio incluido proporcionó datos sobre este desenlace.

Una desviación respecto al protocolo afectó a la caracterización de la exposición. Aunque inicialmente se habían establecido criterios más restrictivos, los autores decidieron evaluar también los estudios en los que el nivel de exposición sólo podía inferirse a partir de las condiciones experimentales asumidas y no había sido determinado directamente mediante medición o estimación. La razón fue la dificultad de establecer un límite inequívoco dentro del continuo de calidad de la información dosimétrica disponible.

Selección y extracción de datos

La selección de los estudios se realizó en dos etapas. Dos revisores evaluaron independientemente los títulos y resúmenes de los registros identificados. Cuando existía desacuerdo o el resumen no proporcionaba información suficiente para determinar la elegibilidad, el artículo pasaba a la evaluación del texto completo. Posteriormente, dos revisores evaluaron de forma independiente los textos completos y las discrepancias se resolvieron mediante discusión o mediante la intervención de un tercer revisor.

Cuando diferentes artículos comunicaban resultados procedentes de un mismo estudio experimental, se consideraron como un único estudio. No fue necesario contactar con los autores para resolver dudas relacionadas con la selección. Los artículos publicados en idiomas distintos del inglés fueron traducidos por los propios revisores o mediante Google Translate.

La extracción de las características de los estudios y de los datos correspondientes a los desenlaces siguió el procedimiento establecido en el protocolo. Se recopiló información sobre el modelo animal, las condiciones experimentales, las características de la exposición, la dosimetría disponible, los grupos comparadores y los datos necesarios para calcular las correspondientes medidas de efecto. Dada la naturaleza de los desenlaces reproductivos y del desarrollo, se prestó especial atención a la identificación de la unidad experimental y a la dependencia de las observaciones procedentes de una misma camada.

Medidas de efecto y síntesis cuantitativa

La síntesis cuantitativa se realizó mediante metaanálisis, seleccionando la medida de efecto en función de la naturaleza y de la forma de expresión de cada desenlace. Se emplearon diferencias de medias (MD), diferencias de medias estandarizadas (SMD) y *odds ratios* (OR), según correspondiera. La SMD permitió combinar resultados expresados mediante escalas o unidades diferentes, mientras que las OR se utilizaron para determinados desenlaces dicotómicos. Los análisis tuvieron en cuenta, cuando resultaba necesario, la falta de independencia de los fetos pertenecientes a una misma camada.

Los estudios clasificados con preocupación baja o intermedia por riesgo de sesgo constituyeron el cuerpo principal utilizado para estimar los efectos y evaluar la certeza de la evidencia. Los resultados de los estudios clasificados como *high concern* se presentaron separadamente, permitiendo comparar su comportamiento con el de los estudios con menor preocupación por riesgo de sesgo. Esta separación resulta relevante porque en algunos desenlaces las estimaciones obtenidas en ambos conjuntos diferían de forma apreciable.

Cuando los experimentos incluían varios niveles de exposición, los autores realizaron análisis que permitían obtener una estimación agregada de los grupos expuestos frente al correspondiente control y, cuando los datos lo permitían, análisis exposición-respuesta utilizando los distintos niveles disponibles. Estos últimos se emplearon para investigar si la magnitud de los efectos variaba sistemáticamente con el nivel de exposición. Se exploraron modelos lineales y, cuando procedía, modelos no lineales, comparando su ajuste para valorar la forma de la relación exposición-respuesta. Por ejemplo, en determinados desenlaces se comparó el ajuste lineal con funciones *spline* cúbicas.

La heterogeneidad entre estudios se evaluó estadísticamente y se exploraron posibles fuentes mediante análisis de subgrupos. Entre los factores considerados se encontraban la especie animal, la SAR y el incremento de temperatura corporal asociado a la exposición. Estos análisis tenían especial relevancia porque permitían valorar si las diferencias entre experimentos podían relacionarse con las condiciones físicas de exposición y, particularmente, con niveles capaces de producir calentamiento apreciable. No todos los análisis de subgrupos previstos pudieron realizarse, especialmente cuando alguna categoría contenía un número insuficiente de estudios.

La posibilidad de sesgo de publicación se examinó mediante gráficos en embudo (*funnel plots*) y, cuando el número de estudios lo permitía, mediante pruebas estadísticas de asimetría. Esta evaluación estuvo necesariamente limitada para aquellos desenlaces representados por un número reducido de experimentos.

Evaluación del riesgo de sesgo (RoB)

El riesgo de sesgo se evaluó siguiendo las directrices OHAT del NTP, adaptadas a las características de los estudios experimentales en animales. La valoración consideró diferentes dominios relacionados con la validez interna de los experimentos, entre ellos la asignación de los animales a los grupos experimentales, el cegamiento, la caracterización de la exposición, la evaluación de los desenlaces, la integridad de los datos y la posible notificación selectiva.

A partir de las valoraciones de los distintos dominios las evaluaciones de cada desenlace se clasificaron en tres categorías de preocupación por riesgo de sesgo: *low concern*, *some concern* y *high concern*. Esta clasificación tuvo consecuencias directas para la síntesis de la evidencia. Los estudios considerados *high concern* se mantuvieron visibles y sus resultados se analizaron separadamente, pero no formaron parte del cuerpo de evidencia utilizado para establecer las conclusiones principales y la certeza mediante GRADE. La propia revisión presenta de forma diferenciada las estimaciones agrupadas correspondientes a los estudios con *low/some concern* y las procedentes de estudios con *high concern*.

La evaluación de la calidad de la exposición fue particularmente importante, dado que los estudios diferían considerablemente en la información dosimétrica disponible. Los autores valoraron la generación y caracterización de la exposición y la disponibilidad y fiabilidad de parámetros como la SAR, así como la información relativa al incremento térmico producido durante los experimentos.

Certeza de la evidencia (GRADE)

La certeza del cuerpo de evidencia se evaluó mediante el enfoque GRADE para cada desenlace, utilizando como base los estudios que no habían sido clasificados como *high concern* por riesgo de sesgo. La valoración consideró los dominios de riesgo de sesgo, inconsistencia, indirectitud, imprecisión y sesgo de publicación. Cuando resultaba aplicable, se valoró asimismo la consistencia de los resultados entre especies como posible factor para incrementar la certeza.

La evaluación se realizó de forma independiente para los diferentes desenlaces relacionados con reducción de la fecundidad, efectos adversos presentes al nacimiento y efectos diferidos sobre la salud de la descendencia. No todos ellos pudieron recibir una valoración equivalente: cuando el número de estudios de calidad suficiente era demasiado pequeño para realizar un metaanálisis, la capacidad para establecer formalmente la certeza quedaba limitada. Por ejemplo, para las pérdidas preimplantacionales sólo existía un estudio no clasificado como *high concern*, por lo que no se realizó metaanálisis para el perfil GRADE de este desenlace.

La indirectitud se valoró atendiendo a la capacidad de los modelos experimentales y de cada desenlace para informar sobre posibles efectos relevantes para la salud humana. Asimismo, la heterogeneidad entre especies, sistemas de exposición y niveles dosimétricos, junto con la precisión de las estimaciones y la posible existencia de sesgo de publicación, condicionaron la valoración final.

Por tanto, la revisión no asignó una certeza única al conjunto de los estudios experimentales, sino que estableció niveles de certeza específicos para cada desenlace. Este procedimiento permite distinguir entre resultados respaldados por un cuerpo de evidencia experimental relativamente consistente y aquellos cuya interpretación queda limitada por riesgo de sesgo, heterogeneidad, indirectitud, imprecisión o un número insuficiente de estudios.

3.4.4. Resultados

Flujo PRISMA

La búsqueda bibliográfica identificó 4.326 registros en PubMed (n = 1.095), Scopus (n = 1.889) y EMF-Portal (n = 1.342). Tras eliminar 1.132 duplicados, se cribaron 3.194 registros mediante la revisión del título y el resumen, de los cuales se excluyeron 2.958.

Se intentó recuperar el texto completo de 236 publicaciones. De estas, 11 no pudieron recuperarse y 10 no pudieron traducirse, por lo que se evaluaron 215 publicaciones para determinar su elegibilidad. En esta fase se excluyeron 127 publicaciones por los siguientes motivos: diseño experimental no realizado *in utero* (n = 34), caracterización insuficiente de la exposición (n = 16), tipo de exposición fuera del alcance de la revisión (n = 12), contraste de exposición insuficiente (n = 4), ausencia de un comparador adecuado (n = 6), datos de evaluación de los desenlaces no válidos o fuera del alcance de la revisión (n = 45), publicación no correspondiente a un artículo científico original y revisado por pares (n = 9) y publicación duplicada (n = 1). Finalmente, se incluyeron 88 artículos en la revisión sistemática.

La unidad considerada como «estudio» en los análisis no coincidió necesariamente con la publicación. Un mismo artículo podía contener varios experimentos independientes, diferentes niveles de exposición o comparaciones con distintos grupos *sham*. Cuando estas comparaciones constituían unidades experimentales independientes, se trataron como estudios separados. Por ello, el número de estudios que contribuyó a determinados metaanálisis pudo ser superior al número de publicaciones correspondientes.

Características de los estudios incluidos

Los 88 artículos incluidos correspondieron exclusivamente a estudios experimentales controlados realizados en mamíferos no humanos expuestos *in utero* a CEM-RF. La evidencia procedía fundamentalmente de roedores, principalmente ratas y ratones, aunque también se incluyeron otras especies de mamíferos cuando cumplían los criterios establecidos. Los experimentos abarcaron diferentes periodos de la gestación, desde exposiciones restringidas a determinadas ventanas del desarrollo embrionario o fetal hasta exposiciones que comprendían prácticamente toda la gestación.

Las condiciones de exposición fueron muy heterogéneas. Se utilizaron diferentes frecuencias, modulaciones, sistemas de exposición, duraciones diarias y periodos gestacionales, así como un amplio intervalo de niveles de exposición. La revisión prestó especial atención a la SAR media de cuerpo entero como parámetro dosimétrico. Cuando esta no estaba comunicada directamente, pero existía información suficiente, los autores intentaron estimarla a partir de las características dosimétricas disponibles. También se registró si la exposición producía un incremento de la temperatura corporal, aspecto particularmente relevante porque los efectos térmicos de exposiciones suficientemente intensas constituyen un mecanismo conocido de toxicidad embriofetal.

Los desenlaces se organizaron en tres grandes categorías. La primera fue la reducción de la fecundidad, que comprendió pérdidas preimplantacionales, tamaño de camada y fetos reabsorbidos o muertos. La segunda correspondió a efectos adversos sobre la salud de la descendencia al nacimiento, incluyendo peso y longitud fetal, malformaciones congénitas y proporción de sexos. La tercera agrupó los efectos tardíos sobre la salud de la descendencia, fundamentalmente alteraciones cerebrales y neuroconductuales, fertilidad de las hembras descendientes y cáncer de aparición temprana. Los desenlaces evaluados al nacimiento se consideraron prioritarios por su relevancia para la salud humana y por la mayor proximidad temporal entre la exposición prenatal y el efecto.

La cantidad de evidencia fue muy diferente según el desenlace. Por ejemplo, el tamaño de camada estuvo representado por 29 artículos que aportaron 47 estudios, el peso fetal por 44 artículos y 104 estudios, y los fetos reabsorbidos o muertos por 16 artículos y 76 estudios. En otros desenlaces, especialmente algunos efectos tardíos, el número de estudios fue considerablemente menor.

Reducción de la fecundidad

La evidencia sobre pérdidas preimplantacionales fue insuficiente para realizar una síntesis cuantitativa robusta. Aunque se identificaron varios trabajos con información relacionada con este desenlace, solamente un estudio que evaluaba pérdidas preimplantacionales fue considerado de riesgo de sesgo bajo o con algunas preocupaciones; por este motivo, el desenlace no fue sometido a evaluación GRADE.

Para el tamaño de camada, el metaanálisis principal incluyó 24 estudios clasificados como de riesgo de sesgo bajo o con algunas preocupaciones. La diferencia media agrupada fue de 0,05 crías por camada (IC95 % -0,21 a 0,30), resultado compatible con ausencia de efecto de la exposición prenatal a CEM-RF. La SAR media de cuerpo entero de los estudios incluidos en este análisis fue de 4,92 W/kg. Los análisis dosis-respuesta tampoco aportaron evidencia de una disminución del tamaño de camada con el aumento de la exposición. La certeza de esta evidencia fue calificada como alta.

El resultado fue diferente para el número de fetos reabsorbidos o muertos. El metaanálisis de 21 estudios de riesgo bajo o con algunas preocupaciones mostró una OR de 1,84 (IC95 % 1,27-2,66), indicativa de una mayor incidencia en los animales expuestos. La SAR media de cuerpo entero fue, sin embargo, considerablemente mayor que en los estudios de tamaño de camada: 20,26 W/kg. Los resultados del análisis dosis-respuesta fueron concordantes con el análisis principal. La certeza se calificó como baja, por lo que los autores consideran que los CEM-RF pueden incrementar la incidencia de fetos reabsorbidos o muertos, pero existe incertidumbre considerable sobre esta estimación.

Cordelli et al. destacan precisamente la aparente discrepancia entre ambos desenlaces. La SAR media de cuerpo entero fue aproximadamente cinco veces mayor en los estudios sobre fetos reabsorbidos o muertos que en los estudios sobre tamaño de camada, y el número de estudios realizados a 5 W/kg o más fue aproximadamente el doble. Por ello, las diferencias entre ambos resultados podrían estar relacionadas, al menos parcialmente, con los niveles de exposición utilizados.

Efectos adversos sobre la descendencia al nacimiento

El peso fetal constituyó el desenlace con una de las bases experimentales más amplias. El metaanálisis de 48 estudios de riesgo bajo o con algunas preocupaciones mostró una SMD de 0,31 (IC95 % 0,15–0,48), donde el signo positivo fue definido por los autores de forma que representase una reducción del peso fetal y, por tanto, un efecto adverso. La SAR media de cuerpo entero fue de 9,83 W/kg. Los autores interpretaron la magnitud como un pequeño efecto perjudicial de la exposición sobre el peso fetal y calificaron la certeza como moderada.

Para la longitud fetal, 13 estudios contribuyeron al metaanálisis principal. La estimación agrupada fue SMD = 0,45 (IC95 % 0,07–0,83), compatible con una reducción moderada de la longitud en los animales expuestos. La SAR media de cuerpo entero fue de 4,55 W/kg. La certeza de la evidencia se calificó como baja, por lo que, aunque la estimación agrupada mostró un efecto perjudicial, existe una incertidumbre considerable sobre este resultado.

Las malformaciones fetales se analizaron mediante dos aproximaciones. Cuando se consideró la fracción o proporción de fetos malformados, 13 estudios proporcionaron una SMD = –0,45 (IC95 % –0,68 a –0,23); en esta métrica, el signo negativo representa un efecto perjudicial. La SAR media de cuerpo entero fue de 6,75 W/kg. Cuando el desenlace se expresó como número de camadas con fetos malformados, 28 estudios produjeron una OR = 3,22 (IC95 % 1,90–5,46), con una SAR media de cuerpo entero considerablemente superior, 16,63 W/kg. Ambos análisis mostraron, por tanto, una mayor frecuencia de malformaciones en los grupos expuestos, aunque la certeza de la evidencia fue baja o muy baja y los niveles de exposición fueron un elemento fundamental para interpretar estos resultados.

Para la proporción de sexos, el metaanálisis de 13 estudios produjo una OR = 1,08 (IC95 % 0,92–1,28), compatible con poca o ninguna diferencia entre los grupos expuestos y *sham*.

En conjunto, los análisis de los desenlaces al nacimiento muestran un patrón diferente según el parámetro considerado: evidencia de alta certeza de ausencia de efecto sobre el tamaño de camada, evidencia de certeza moderada de una pequeña reducción del peso fetal y señales de efectos adversos sobre longitud fetal, reabsorciones/muerte fetal y malformaciones, pero con menor certeza y, en varios casos, en experimentos caracterizados por SAR medias relativamente elevadas.

Efectos tardíos sobre la descendencia

Los efectos tardíos estuvieron respaldados por una base experimental considerablemente menor y, en especial para los desenlaces neuroconductuales, por una escasa replicación independiente entre laboratorios.

Para la patología cerebral, representada principalmente por el peso cerebral, 12 estudios pudieron combinarse y la estimación agrupada fue SMD = 0,10 (IC95 % –0,09 a 0,29), compatible con ausencia de un efecto adverso relevante. Los autores calificaron como moderada la certeza de la evidencia de ausencia de efecto sobre el peso cerebral.

Para aprendizaje y memoria, únicamente dos estudios de riesgo bajo o con algunas preocupaciones contribuyeron al metaanálisis principal, con una SMD = –0,54 (IC95 % –1,24 a 0,17).

El intervalo de confianza incluyó el valor nulo y, por tanto, no permitió establecer un efecto perjudicial. Resulta especialmente importante que buena parte de la literatura disponible para este desenlace se encontrara en estudios clasificados como de alta preocupación por riesgo de sesgo: al analizarlos separadamente, nueve estudios proporcionaban una SMD de $-0,57$ (IC95 % $-1,00$ a $-0,14$). Esta discrepancia ilustra la influencia de la calidad metodológica sobre la síntesis.

Para actividad motora, cuatro estudios de riesgo bajo o con algunas preocupaciones mostraron una SMD = $0,79$ (IC95 % $0,21$ – $1,38$), interpretada como un efecto perjudicial grande. Para las funciones motoras y sensoriales, dos estudios proporcionaron una SMD = $-0,66$ (IC95 % $-1,18$ a $-0,14$), compatible con un efecto perjudicial moderado. Sin embargo, ambos resultados se apoyaron en pocos estudios y con limitada replicación independiente, por lo que la certeza fue muy baja.

La fertilidad de las hembras descendientes se evaluó mediante el tamaño de las camadas producidas posteriormente por las hembras F2. Cuatro estudios mostraron una SMD = $0,08$ (IC95 % $-0,39$ a $0,55$), compatible con poca o ninguna diferencia entre expuestas y controles. La certeza se calificó como baja.

En conjunto, los autores consideran que los CEM-RF probablemente no afectan al peso cerebral de la descendencia y pueden no reducir la fertilidad femenina posterior. Por el contrario, algunos resultados apuntan a posibles efectos adversos neuroconductuales, pero la escasez de estudios, la limitada replicación independiente y la muy baja certeza impiden establecer conclusiones firmes.

Influencia del nivel de exposición, SAR y temperatura

La interpretación de los resultados está condicionada de manera especialmente importante por la dosimetría. Los niveles medios de SAR variaron ampliamente entre desenlaces: desde aproximadamente $4,5$ – 5 W/kg para longitud fetal y tamaño de camada hasta $16,63$ W/kg para camadas con fetos malformados y $20,26$ W/kg para fetos reabsorbidos o muertos. Algunos de estos niveles son suficientemente elevados como para que la contribución térmica constituya un elemento relevante de interpretación.

Los autores realizaron análisis de subgrupos según especie, SAR ($<0,1$; $0,1$ a <5 ; ≥ 5 W/kg) e incremento de la temperatura corporal (<1 °C frente a ≥ 1 °C), además de análisis dosis-respuesta lineales y no lineales. Estos análisis permitieron explicar parcialmente la heterogeneidad para algunos desenlaces, pero no resolvieron de forma general la incertidumbre acerca de los efectos a niveles bajos de exposición. De hecho, una de las conclusiones metodológicas centrales de la revisión es que el cuerpo de evidencia disponible no permite determinar si los CEM-RF afectan a los resultados del embarazo a niveles de exposición inferiores a los capaces de producir el conocido efecto adverso del calentamiento.

Sesgo de publicación

Los autores evaluaron el posible sesgo de publicación mediante *funnel plots* y la prueba de Egger utilizando todos los estudios recuperados, independientemente de su clasificación de riesgo de sesgo. Se encontraron indicios de posible sesgo de publicación para las malformaciones, tanto cuando se analizaron como variable continua como binaria, con resultados limítrofes de la prueba de Egger.

También se observaron indicios para aprendizaje y memoria a partir de la inspección del *funnel plot* y de la prueba de Egger. Para los restantes desenlaces no se identificaron indicios de sesgo de publicación.

Evaluación del riesgo de sesgo

El riesgo de sesgo se evaluó mediante una adaptación de la herramienta OHAT para estudios experimentales en animales. Los autores clasificaron cada evaluación de desenlace como de baja preocupación (*low concern*), algunas preocupaciones (*some concern*) o alta preocupación (*high concern*). La evaluación se realizó a nivel de desenlace, de modo que una misma publicación podía recibir valoraciones diferentes para distintos resultados.

Entre los aspectos considerados figuraron la aleatorización, ocultación de la asignación, cegamiento durante la exposición y la evaluación de los desenlaces, adecuación de los controles *sham*, pérdidas o exclusiones, notificación selectiva y calidad de la caracterización de la exposición. Se incorporó específicamente una cuestión relativa a si el posible incremento de temperatura inducido por CEM-RF había sido evaluado adecuadamente, dada su relevancia para distinguir los efectos de la radiofrecuencia de los derivados del calentamiento.

Los estudios clasificados como de alta preocupación se mostraron separadamente en los *forest plots*, pero fueron excluidos de los análisis posteriores utilizados para establecer las conclusiones principales y de la evaluación de certeza mediante GRADE. Esta decisión tuvo consecuencias relevantes para algunos desenlaces. Por ejemplo, para tamaño de camada, los 24 estudios de riesgo bajo o con algunas preocupaciones mostraron MD = 0,05 (IC95 % -0,21 a 0,30), mientras que los 12 estudios de alta preocupación mostraron MD = 0,77 (IC95 % 0,15-1,39). En aprendizaje y memoria, los dos estudios de mayor calidad mostraron SMD = -0,54 (IC95 % -1,24 a 0,17), mientras que nueve estudios de alta preocupación produjeron SMD = -0,57 (IC95 % -1,00 a -0,14). Estos ejemplos muestran que la consideración explícita del riesgo de sesgo modificaba en algunos casos la interpretación de la evidencia.

Certeza de la evidencia (GRADE)

La certeza de la evidencia se evaluó mediante GRADE utilizando exclusivamente los estudios clasificados como de riesgo de sesgo bajo o con algunas preocupaciones. La valoración consideró riesgo de sesgo, inconsistencia, indirectitud, imprecisión y posible sesgo de publicación, además de la posibilidad de aumentar la certeza cuando existía consistencia entre especies.

La conclusión con mayor grado de certeza correspondió al tamaño de camada: la evidencia fue calificada como de certeza alta y mostró que la exposición prenatal a CEM-RF no afecta al tamaño de la camada (MD = 0,05 crías; IC95 % -0,21 a 0,30; 24 estudios). Para el peso fetal, la evidencia de una pequeña reducción asociada a la exposición se calificó como de certeza moderada. También fue moderada la certeza de la evidencia de ausencia de un efecto tardío sobre el peso cerebral de la descendencia.

La evidencia sobre el aumento de fetos reabsorbidos o muertos fue de certeza baja (OR = 1,84; IC95 % 1,27–2,66; 21 estudios). También fue baja la certeza respecto a la ausencia o escaso efecto sobre la fertilidad posterior de las hembras descendientes (SMD = 0,08; IC95 % –0,39 a 0,55). Para varios de los restantes desenlaces, incluidos algunos resultados de malformaciones y neuroconductuales, la certeza fue baja o muy baja, debido a combinaciones de riesgo de sesgo, inconsistencia, imprecisión, posible sesgo de publicación y escasez de replicación independiente.

Por tanto, la revisión no ofrece una conclusión uniforme de «efecto» o «ausencia de efecto» de los CEM-RF sobre el desarrollo prenatal. La evidencia de mayor certeza indica ausencia de efecto sobre el tamaño de camada, mientras que existe evidencia de certeza moderada de una pequeña reducción del peso fetal. Para otros desenlaces se observan asociaciones adversas, algunas de magnitud apreciable, pero sustentadas por evidencia de certeza baja o muy baja. Además, los elevados niveles de SAR empleados en una parte importante de la literatura impiden establecer a partir de este cuerpo experimental si esos efectos se producen a exposiciones inferiores a las capaces de generar efectos térmicos adversos.

3.4.5. Fortalezas y debilidades de la revisión

Fortalezas

Una de las principales fortalezas de la revisión de Cordelli et al. (2023) es su carácter sistemático y metodológicamente predefinido. El trabajo se realizó de acuerdo con el protocolo publicado previamente por Pacchierotti et al. (2021) y registrado en PROSPERO (CRD42021227746), siguiendo las directrices de la OMS y del *National Toxicology Program/Office of Health Assessment and Translation* (NTP/OHAT). Los autores especificaron previamente la pregunta PECO, los criterios de elegibilidad, los desenlaces de interés, el procedimiento de evaluación del riesgo de sesgo y la estrategia general de síntesis. Además, las desviaciones respecto al protocolo fueron identificadas explícitamente en el artículo, lo que aumenta la transparencia metodológica.

La búsqueda bibliográfica fue amplia y no impuso restricciones por año de publicación ni por idioma. Se utilizaron PubMed y Scopus junto con EMF-Portal, específicamente orientado a la literatura sobre campos electromagnéticos. La selección se realizó independientemente por dos revisores y los desacuerdos se resolvieron mediante discusión o intervención de un tercero. La extracción de las características y resultados fue realizada por un revisor y comprobada íntegramente por un segundo; cuando faltaban datos esenciales o existían inconsistencias, los autores de los estudios primarios fueron contactados. Estas medidas reducen el riesgo de errores durante la selección y extracción de la evidencia.

Otra fortaleza importante es la definición estricta de las condiciones experimentales elegibles. La revisión se limitó a estudios controlados en mamíferos expuestos exclusivamente *in utero* y exigió controles *sham* adecuados o controles sometidos al mismo incremento térmico mediante calentamiento directo.

Además, los autores prestaron especial atención a la dosimetría: cuando no se proporcionaba directamente la SAR media de cuerpo entero, intentaron estimarla a partir de la información dosimétrica disponible. Esta caracterización permitió no sólo realizar metaanálisis convencionales, sino también explorar relaciones dosis-respuesta y efectuar análisis de subgrupos según nivel de SAR y aumento de la temperatura corporal.

La selección de los desenlaces constituye asimismo una fortaleza. Se priorizaron parámetros con relevancia directa para la salud y se tuvieron en cuenta las guías de ensayo de la OCDE sobre toxicidad del desarrollo prenatal y neurotoxicidad del desarrollo (OECD TG 414 y 426). Los resultados se organizaron en reducción de la fecundidad, efectos adversos sobre la descendencia al nacimiento y efectos tardíos sobre la salud de la descendencia, evitando en lo posible la utilización de alteraciones moleculares cuya relación causal con resultados adversos para la salud no estuviera establecida.

También resulta especialmente valioso el tratamiento explícito del riesgo de sesgo. La herramienta OHAT fue adaptada a las particularidades de los estudios experimentales con CEM-RF, incorporando específicamente la cuestión de si el posible incremento de temperatura inducido por la exposición había sido adecuadamente considerado y evaluado. Los estudios clasificados como de *high concern* se presentaron separadamente, pero fueron excluidos de los análisis empleados para establecer las conclusiones principales y valorar la certeza mediante GRADE. De este modo, las conclusiones no dependen indiscriminadamente de toda la literatura publicada, sino preferentemente del subconjunto experimental considerado metodológicamente más fiable.

Finalmente, la revisión no se limita a establecer si existe una diferencia estadística entre animales expuestos y controles. El empleo de modelos de efectos aleatorios, análisis de subgrupos por especie, SAR y aumento de temperatura, metaanálisis dosis-respuesta lineales y no lineales, evaluación del posible sesgo de publicación y valoración formal de la certeza mediante GRADE proporciona una evaluación mucho más completa de la evidencia. Esta aproximación resulta particularmente importante en este campo, porque permite analizar hasta qué punto los efectos observados dependen de exposiciones elevadas y potencialmente térmicas, cuestión esencial para interpretar su relevancia respecto de exposiciones humanas habituales.

Limitaciones

La principal limitación de la revisión no reside tanto en su diseño como en las características del cuerpo de evidencia disponible. Los 88 artículos incluidos presentaban una considerable heterogeneidad en especies y cepas animales, sistemas de exposición, frecuencias y modulaciones, periodos gestacionales expuestos, duración de la exposición, niveles de SAR y procedimientos de evaluación de los desenlaces. Aunque los modelos de efectos aleatorios y los análisis de subgrupos permiten abordar parcialmente esta variabilidad, no eliminan la dificultad de integrar experimentos biológica y técnicamente muy diferentes.

Una limitación especialmente relevante es la calidad de la caracterización de la exposición en una parte de los estudios primarios. No todos proporcionaron directamente la SAR media de cuerpo entero y, en algunos casos, fue necesario estimarla a partir de otras magnitudes y supuestos biofísicos.

Además, la revisión terminó incluyendo estudios en los que el nivel de exposición sólo podía inferirse a partir de las condiciones experimentales, en lugar de disponer de una medición o estimación dosimétrica propiamente dicha. Los autores reconocen explícitamente esta modificación respecto al protocolo, justificándola por la dificultad de establecer un umbral objetivo dentro del continuo de calidad de la información dosimétrica.

La dosimetría condiciona además la interpretación biológica de los resultados. Algunos de los efectos adversos se observaron en conjuntos de experimentos realizados con SAR medias de cuerpo entero elevadas: 9,83 W/kg para peso fetal, 16,63 W/kg para camadas con fetos malformados y 20,26 W/kg para fetos reabsorbidos o muertos.

Dado que el calentamiento inducido por CEM-RF a intensidades suficientemente altas es un efecto adverso conocido y que el desarrollo embrionario y fetal es particularmente sensible a la temperatura, resulta difícil separar los posibles efectos no térmicos de aquellos mediados por el incremento de temperatura. Esta es una limitación central de la base experimental, y los propios autores concluyen que el cuerpo de evidencia no permite determinar si los CEM-RF afectan a los resultados del embarazo a niveles inferiores a los que producen el conocido efecto adverso del calentamiento.

El riesgo de sesgo de los estudios primarios constituye otra limitación importante. La información sobre aleatorización, ocultación de la asignación y cegamiento durante la exposición o la evaluación de los desenlaces fue incompleta o inadecuada en una parte de la literatura. También existieron problemas relacionados con la calidad de los controles *sham*, la caracterización de la exposición, las exclusiones de animales y la evaluación del posible aumento de temperatura. La decisión de excluir de los análisis principales los estudios clasificados como de *high concern* reduce la influencia de estos problemas sobre las conclusiones, pero al mismo tiempo disminuye el número de experimentos disponibles para algunos desenlaces.

Existen además limitaciones estadísticas derivadas de la estructura de los estudios de toxicidad reproductiva. La unidad experimental apropiada es la madre o camada y no cada feto o cría individual. Cuando los datos originales no estaban claramente expresados como medias por camada, los autores aplicaron una corrección por agrupamiento intracamadas utilizando un coeficiente de correlación intraclase asumido de 0,2.

Para numerosos estudios sobre efectos tardíos ni siquiera estaba disponible el número de madres, por lo que esta corrección no pudo aplicarse. Los propios autores reconocen que ello puede introducir un sesgo hacia la detección de efectos de los CEM-RF, aunque consideran que probablemente tendría una influencia pequeña en los resultados tardíos.

La evidencia sobre determinados efectos tardíos, especialmente los neuroconductuales, es además escasa y presenta una limitada replicación independiente. Algunos metaanálisis se apoyan en sólo dos o cuatro estudios clasificados como de riesgo bajo o con algunas preocupaciones y, en determinados casos, estos proceden de pocos laboratorios. Por ello, asociaciones aparentemente importantes, como las observadas para actividad motora o funciones motoras y sensoriales, presentan una certeza muy baja y no pueden considerarse resultados consolidados. La ausencia de replicación independiente constituye aquí una limitación diferente de la mera imprecisión estadística, porque dificulta saber si los resultados son reproducibles entre grupos experimentales distintos.

Por último, la traslación de los resultados de estudios experimentales en mamíferos a la exposición humana requiere cautela. Los modelos animales son esenciales para estudiar de forma controlada la toxicidad prenatal y permiten exposiciones y procedimientos imposibles en humanos, pero las condiciones experimentales, y especialmente algunas de las SAR empleadas, no representan necesariamente las exposiciones habituales de la población.

En consecuencia, la revisión es informativa sobre la existencia potencial de efectos biológicos y sobre su relación con la dosis y el calentamiento, pero no permite trasladar directamente todos los tamaños de efecto observados a situaciones reales de exposición humana. Esta limitación de aplicabilidad resulta particularmente relevante para los desenlaces en los que las asociaciones adversas proceden fundamentalmente de experimentos realizados a niveles elevados de exposición.

En conjunto, Cordelli et al. realizan una revisión sistemática metodológicamente rigurosa, con evaluación explícita del riesgo de sesgo, dosimetría, temperatura, relaciones dosis-respuesta y certeza de la evidencia. Sin embargo, su capacidad para responder a la cuestión de mayor relevancia para la evaluación del riesgo, si la exposición prenatal a CEM-RF produce efectos adversos a niveles inferiores a aquellos capaces de generar efectos térmicos, está limitada por las características de los estudios primarios disponibles. Esta distinción es esencial: la incertidumbre restante no deriva simplemente de que existan resultados contradictorios, sino de la combinación de calidad metodológica variable, heterogeneidad experimental, escasa replicación para determinados desenlaces y, sobre todo, de la utilización frecuente de niveles de exposición elevados que dificultan separar los posibles efectos de la RF de los efectos térmicos conocidos.

3.4.6. Conclusiones

En conjunto, Cordelli et al. (2023) concluyen que la evidencia experimental disponible sobre exposición prenatal a CEM-RF no permite formular una conclusión única aplicable a todos los resultados reproductivos y del desarrollo. Los efectos observados difieren según el desenlace considerado, y su interpretación está condicionada tanto por la certeza de la evidencia como por los niveles de exposición empleados en los experimentos. Un aspecto central de la revisión es que una parte importante de los resultados adversos procede de estudios realizados con SAR de cuerpo entero relativamente elevadas, lo que dificulta determinar hasta qué punto los efectos observados se deben a mecanismos distintos del calentamiento o al efecto térmico inducido por exposiciones suficientemente intensas a CEM-RF.

Para la reducción de la fecundidad, la evidencia de mayor certeza corresponde al tamaño de camada. El metaanálisis de los estudios clasificados como de riesgo de sesgo bajo o con algunas preocupaciones mostró una diferencia media de 0,05 crías por camada (IC95 % -0,21 a 0,30), compatible con ausencia de efecto. La certeza de esta evidencia fue alta, por lo que los autores concluyen que la exposición prenatal a CEM-RF no afecta al tamaño de la camada en las condiciones estudiadas. En cambio, para los fetos reabsorbidos o muertos se observó un incremento en los grupos expuestos (OR = 1,84; IC95 % 1,27-2,66), aunque con certeza baja. Esta diferencia entre ambos desenlaces debe interpretarse teniendo en cuenta que los experimentos sobre reabsorción o muerte fetal utilizaron niveles de exposición considerablemente mayores: la SAR media de cuerpo entero fue de 20,26 W/kg, frente a 4,92 W/kg en los estudios sobre tamaño de camada.

Entre los efectos adversos sobre la descendencia al nacimiento, la evidencia más consistente de un efecto perjudicial correspondió al peso fetal. El metaanálisis mostró una pequeña reducción del peso en los animales expuestos ($SMD = 0,31$; $IC95\% 0,15-0,48$, con la dirección de la escala definida de manera que los valores positivos representan un efecto adverso), con una certeza moderada.

También se observaron asociaciones compatibles con una reducción de la longitud fetal y con una mayor frecuencia de malformaciones. Sin embargo, la certeza para estos desenlaces fue inferior y, nuevamente, algunos de los niveles de exposición fueron elevados. En particular, la SAR media de cuerpo entero alcanzó $16,63\text{ W/kg}$ en el análisis de camadas con fetos malformados.

Los efectos tardíos sobre la descendencia presentan un panorama todavía más incierto. La evidencia de certeza moderada indica que la exposición prenatal probablemente no produce un efecto relevante sobre el peso cerebral. Asimismo, los datos disponibles sugieren poca o ninguna alteración posterior de la fertilidad de las hembras descendientes, aunque en este caso la certeza fue baja.

Para diferentes resultados neuroconductuales se observaron asociaciones compatibles con efectos adversos, particularmente en actividad motora y funciones motoras y sensoriales, pero la evidencia fue de certeza muy baja, se basó en pocos estudios y presentó una limitada replicación independiente. Por ello, estos resultados no permiten establecer conclusiones firmes sobre efectos neuroconductuales tardíos de la exposición prenatal.

La dosimetría resulta fundamental para interpretar el conjunto de estos hallazgos. Los análisis muestran que los experimentos que detectaron determinados efectos adversos utilizaron con frecuencia SAR superiores a las empleadas en los estudios que no encontraron efectos. Además, el desarrollo embrionario y fetal es especialmente sensible al aumento de temperatura, y el calentamiento producido por exposiciones suficientemente intensas a RF constituye un mecanismo conocido de toxicidad. Por consiguiente, la presencia de asociaciones adversas en algunos desenlaces no demuestra que esos mismos efectos se produzcan a niveles de exposición inferiores a los capaces de generar un incremento térmico relevante.

Esta consideración conduce a una de las conclusiones más importantes de Cordelli et al.: la evidencia experimental disponible no permite determinar si la exposición prenatal a CEM-RF produce efectos adversos sobre el embarazo o el desarrollo a niveles inferiores a aquellos capaces de producir los efectos adversos conocidos del calentamiento. Por tanto, los resultados de la revisión no deben interpretarse como demostración general de toxicidad reproductiva de los CEM-RF con independencia de la dosis. La evidencia de mayor certeza muestra ausencia de efecto para algunos desenlaces, como el tamaño de camada, un pequeño efecto adverso sobre otros, como el peso fetal, y resultados de menor certeza para reabsorciones, malformaciones y determinados parámetros neuroconductuales.

Desde la perspectiva de la evaluación del riesgo para la salud humana, debe añadirse que las condiciones de muchos de los experimentos no son directamente representativas de las exposiciones habituales de la población. Los estudios animales permiten establecer condiciones controladas y explorar niveles de exposición que no podrían utilizarse experimentalmente en humanos, pero ello no implica que los tamaños de efecto observados puedan extrapolarse directamente a situaciones reales de exposición humana. En consecuencia, la principal incertidumbre relevante no es únicamente si existen efectos en animales, sino a qué niveles de exposición aparecen, hasta qué punto dependen del calentamiento y cuál es su relevancia para las exposiciones humanas.

Implicaciones para investigación futura

Cordelli et al. identifican como prioridad fundamental mejorar la caracterización dosimétrica de los experimentos. Los futuros estudios deberían proporcionar estimaciones fiables de la SAR, describir con precisión la distribución espacial de la exposición y documentar adecuadamente las características del sistema utilizado, incluyendo frecuencia, modulación, duración y calendario de exposición. Una dosimetría insuficiente limita no sólo la reproducibilidad del experimento, sino también la posibilidad de situar sus resultados dentro de una relación exposición-respuesta y compararlos con otros estudios.

Resulta igualmente esencial controlar de manera rigurosa la temperatura. Los experimentos deberían determinar si la exposición produce incrementos térmicos en los animales y disponer de condiciones experimentales que permitan diferenciar los efectos atribuibles a los CEM-RF de los producidos por calentamiento.

Esta cuestión adquiere especial importancia en los estudios con SAR elevadas, precisamente aquellos en los que se observaron algunos de los efectos adversos más pronunciados. Para avanzar en la evaluación del riesgo es necesario generar evidencia específicamente informativa sobre niveles de exposición situados por debajo de aquellos capaces de producir efectos térmicos adversos.

Los autores señalan asimismo la necesidad de utilizar varios niveles de exposición dentro del mismo experimento. Una proporción importante de la literatura compara únicamente un grupo expuesto con un grupo *sham*, lo que limita considerablemente la capacidad para establecer relaciones exposición-respuesta. Los diseños con múltiples niveles de SAR, adecuadamente espaciados y caracterizados dosimétricamente, permitirían determinar con mayor precisión la forma de la relación dosis-respuesta y examinar adecuadamente la relación exposición-respuesta, incluyendo varios niveles de exposición bajo las mismas condiciones experimentales, desde niveles bajos comparables con la exposición humana hasta niveles más elevados en los que puedan esperarse efectos hipertérmicos leves..

También es necesario mejorar aspectos generales del diseño y la notificación experimental. La aleatorización de los animales, el ocultamiento de la asignación, el cegamiento durante la evaluación de los desenlaces, la utilización de controles *sham* apropiados y la descripción completa de las exclusiones y pérdidas deberían incorporarse sistemáticamente.

La revisión muestra que la clasificación por riesgo de sesgo puede modificar sustancialmente algunos resultados agregados, por lo que incrementar la calidad metodológica de los estudios primarios es imprescindible para reducir la incertidumbre.

En los estudios de toxicidad reproductiva y del desarrollo debe prestarse además especial atención a la camada como unidad experimental. Los futuros trabajos deberían informar de forma que permita distinguir claramente el número de madres, camadas, fetos y crías, y utilizar análisis estadísticos que tengan en cuenta la correlación intracamadas. Esto evitaría la pseudorreplicación y reduciría la necesidad de aplicar posteriormente correcciones basadas en coeficientes de correlación asumidos.

Para los efectos tardíos y neuroconductuales, la prioridad no consiste únicamente en aumentar el número de experimentos, sino en conseguir replicaciones independientes de los resultados existentes. Algunos de los hallazgos proceden de pocos estudios o grupos de investigación y presentan certeza muy baja. En este sentido, los autores recomiendan una mejor caracterización de la exposición y de los posibles incrementos de temperatura, junto con una descripción experimental suficientemente estandarizada que facilite la comparación entre estudios.

Finalmente, los futuros estudios deberían orientarse hacia desenlaces con una relación clara con efectos adversos sobre la salud y utilizar procedimientos estandarizados, preferentemente basados en guías reconocidas de toxicología reproductiva y del desarrollo. En este sentido, la revisión pone de manifiesto que el avance del conocimiento no depende simplemente de acumular nuevos experimentos, sino de producir estudios capaces de responder a la cuestión que permanece sin resolver: si existen efectos adversos de la exposición prenatal a CEM-RF por debajo de los niveles en los que el calentamiento constituye una explicación plausible, y cuál sería, en ese caso, su relevancia para la salud humana.

Bibliografía

- Cordelli, E., Ardoino, L., Benassi, B., Consales, C., Eleuteri, P., Marino, C., Sciortino, M., Villani, P., Brinkworth, M.H., Chen, G., Mcnamee, J.P., Wood, A.W., Belackova, L., Verbeek, J., Pacchierotti, F., 2024. Effects of radiofrequency electromagnetic field (RF-EMF) exposure on male fertility: A systematic review of experimental studies on non-human mammals and human sperm *in vitro*. *Environ. Int.* 185, 108509. <https://doi.org/10.1016/j.envint.2024.108509>
- Cordelli, E., Ardoino, L., Benassi, B., Consales, C., Eleuteri, P., Marino, C., Sciortino, M., Villani, P., Brinkworth, M.H., Chen, G., Mcnamee, J.P., Wood, A.W., Belackova, L., Verbeek, J., Pacchierotti, F., 2023. Effects of Radiofrequency Electromagnetic Field (RF-EMF) exposure on pregnancy and birth outcomes: A systematic review of experimental studies on non-human mammals. *Environ. Int.* 180, 108178. <https://doi.org/10.1016/j.envint.2023.108178>
- Johnson, E.E., Kenny, R.P.W., Adesanya, A.M., Richmond, C., Beyer, F., Calderon, C., Rankin, J., Pearce, M.S., Toledano, M., Craig, D., Pearson, F., 2024. The effects of radiofrequency exposure on adverse female reproductive outcomes: A systematic review of human observational studies with dose-response meta-analysis. *Environ Int* 190, 108816. <https://doi.org/10.1016/j.envint.2024.108816>
- Kenny, P.R., Evelynne Johnson, E., Adesanya, A.M., Richmond, C., Beyer, F., Calderon, C., Rankin, J., Pearce, M.S., Toledano, M., Craig, D., Pearson, F., 2024. The effects of radiofrequency exposure on male fertility: A systematic review of human observational studies with dose-response meta-analysis. *Environ Int* 190, 108817. <https://doi.org/10.1016/j.envint.2024.108817>
- Kenny, R.P.W., Millar, E.B., Adesanya, A., Richmond, C., Beyer, F., Calderon, C., Rankin, J., Toledano, M., Feychting, M., Pearce, M.S., Craig, D., Pearson, F., 2022. The effects of radiofrequency exposure on male fertility and adverse reproductive outcomes: A protocol for two systematic reviews of human observational studies with meta-analysis. *Environ. Int.* 158, 106968. <https://doi.org/10.1016/j.envint.2021.106968>
- Pacchierotti, F., Ardoino, L., Benassi, B., Consales, C., Cordelli, E., Eleuteri, P., Marino, C., Sciortino, M., Brinkworth, M.H., Chen, G., McNamee, J.P., Wood, A.W., Hooijmans, C.R., de Vries, R.B.M., 2021. Effects of Radiofrequency Electromagnetic Field (RF-EMF) exposure on male fertility and pregnancy and birth outcomes: Protocols for a systematic review of experimental studies in non-human mammals and in human sperm exposed in vitro. *Environ. Int.* 157, 106806. <https://doi.org/10.1016/j.envint.2021.106806>
- Verbeek, J., Oftedal, G., Feychting, M., van Rongen, E., Rosaria Scarfi, M., Mann, S., Wong, R., van Deventer, E., 2021. Prioritizing health outcomes when assessing the effects of exposure to radiofrequency electromagnetic fields: A survey among experts. *Environment International* 146, 106300. <https://doi.org/10.1016/j.envint.2020.106300>