Environment International 180 (2023) 108178

Contents lists available at ScienceDirect

Environment International

journal homepage: www.elsevier.com/locate/envint

Effects of Radiofrequency Electromagnetic Field (RF-EMF) exposure on pregnancy and birth outcomes: A systematic review of experimental studies on non-human mammals

Eugenia Cordelli [a,*], Lucia Ardoino [a], Barbara Benassi [a], Claudia Consales [a], Patrizia Eleuteri [a,1], Carmela Marino [a,1], Maurizio Sciortino [b,1], Paola Villani [a], Martin H. Brinkworth [c], Guangdi Chen [d], James P. McNamee [e], Andrew W. Wood [f,1], Lea Belackova [g], Jos Verbeek [g], Francesca Pacchierotti [a,*,1]

[a] Division Health Protection Technologies, Italian National Agency for New Technologies, Energy and Sustainable Economic Development (ENEA), Rome, Italy
[b] Department for Sustainability, ENEA, Rome, Italy
[c] School of Chemistry and Bioscience, Faculty of Life Sciences, University of Bradford, Bradford, UK
[d] Bioelectromagnetics Laboratory, Zhejiang University School of Medicine, Hangzhou, China
[e] Non-Ionizing Radiation Health Sciences Division, Consumer and Clinical Radiation Protection Bureau, Health Canada, Ottawa, Canada
[f] Department of Health Sciences and Biostatistics, Swinburne University of Technology, Hawthorn, Australia
[g] University Medical Centers Amsterdam, Coronel Institute of Occupational Health, Cochrane Work, Amsterdam, The Netherlands

ARTICLE INFO

Handling Editor: Paul Whaley

Keywords:
Radiofrequency electromagnetic fields
Adverse pregnancy outcomes
Litter size
Animal studies
Systematic review
Meta-analysis

ABSTRACT

Background: The World Health Organization is coordinating an international project aimed at systematically reviewing the evidence regarding the association between radiofrequency electromagnetic field (RF-EMF) exposure and adverse health effects. Within the project, 6 topics have been prioritized by an expert group, which include reproductive health outcomes.

Objectives: According to the protocol published in 2021, a systematic review and *meta*-analyses on the adverse effects of RF-EMF exposure during pregnancy in offspring of experimental animals were conducted.

Methods: Three electronic databases (PubMed, Scopus and EMF Portal) were last searched on September 8 or 17, 2022. Based on predefined selection criteria, the obtained references were screened by two independent reviewers. Studies were included if they met the following criteria: 1) original, sham controlled experimental study on non-human mammals exposed *in utero*, published in peer-reviewed journals, 2) the experimental RF-EMF exposure was within the frequency range 100 kHz–300 GHz, 3) the effects of RF-EMF exposure on fecundity (litter size, embryonic/fetal losses), on the offspring health at birth (decrease of weight or length, congenital malformations, changes of sex ratio) or on delayed effects (neurocognitive alterations, female infertility or early-onset cancer) were studied. Study characteristics and outcome data were extracted by two reviewers. Risk of bias (RoB) was assessed using the Office of Health Assessment and Translation (OHAT) guidelines. Study results were pooled in a random effects *meta*-analysis comparing average exposure to no-exposure and in a dose–response *meta*-analysis using all exposure doses, after exclusion of studies that were rated at "high concern" for RoB. Subgroup analyses were conducted for species, Specific Absorption Rate (SAR) and temperature increase. The certainty of the evidence was assessed using the Grading of Recommendations, Assessment, Development and Evaluations (GRADE) approach.

Results: Eighty-eight papers could be included in this review. *Effects on fecundity.* The *meta*-analysis of studies on litter size, conducted at a whole-body average SAR of 4.92 W/kg, did not show an effect of RF-EMF exposure (MD 0.05; 95% CI −0.21 to 0.30). The *meta*-analysis of studies on resorbed and dead fetuses, conducted at a whole-body average SAR of 20.26 W/kg, showed a significant increase of the incidence in RF-EMF exposed animals (OR 1.84; 95% CI 1.27 to 2.66). The results were similar in the dose–response analysis. *Effects on the offspring health at birth.* The *meta*-analysis of studies on fetal weight, conducted at a whole-body average SAR of 9.83 W/kg, showed

* Corresponding authors.
E-mail addresses: eugenia.cordelli@enea.it (E. Cordelli), francesca.pacchierotti@gmail.com (F. Pacchierotti).
[1] Retired.

https://doi.org/10.1016/j.envint.2023.108178
Received 13 February 2023; Received in revised form 24 August 2023; Accepted 29 August 2023
Available online 30 August 2023

a small decrease in RF-EMF exposed animals (SMD 0.31; 95% CI 0.15 to 0.48). The *meta*-analysis of studies on fetal length, conducted at a whole-body average SAR of 4.55 W/kg, showed a moderate decrease in length at birth (SMD 0.45; 95% CI 0.07 to 0.83). The *meta*-analysis of studies on the percentage of fetuses with malformations, conducted at a whole-body average SAR of 6.75 W/kg, showed a moderate increase in RF-EMF exposed animals (SMD −0.45; 95% CI −0.68 to −0.23). The *meta*-analysis of studies on the incidence of litters with malformed fetuses, conducted at a whole-body average SAR of 16.63 W/kg, showed a statistically significant detrimental RF-EMF effect (OR 3.22; 95% CI 1.9 to 5.46). The results were similar in the dose–response analyses.
Delayed effects on the offspring health. RF-EMF exposure was not associated with detrimental effects on brain weight (SMD 0.10; 95% CI −0.09 to 0.29) and on learning and memory functions (SMD −0.54; 95% CI −1.24 to 0.17). RF-EMF exposure was associated with a large detrimental effect on motor activity functions (SMD 0.79; 95% CI 0.21 to 1.38) and a moderate detrimental effect on motor and sensory functions (SMD −0.66; 95% CI −1.18 to −0.14). RF-EMF exposure was not associated with a decrease of the size of litters conceived by F2 female offspring (SMD 0.08; 95% CI −0.39 to 0.55). Notably, *meta*-analyses of neurobehavioural effects were based on few studies, which suffered of lack of independent replication deriving from only few laboratories.
Discussion: There was high certainty in the evidence for a lack of association of RF-EMF exposure with litter size. We attributed a moderate certainty to the evidence of a small detrimental effect on fetal weight. We also attributed a moderate certainty to the evidence of a lack of delayed effects on the offspring brain weight. For most of the other endpoints assessed by the *meta*-analyses, detrimental RF-EMF effects were shown, however the evidence was attributed a low or very low certainty. The body of evidence had limitations that did not allow an assessment of whether RF-EMF may affect pregnancy outcomes at exposure levels below those eliciting a well-known adverse heating impact.
In conclusion, *in utero* RF-EMF exposure does not have a detrimental effect on fecundity and likely affects offspring health at birth, based on the *meta*-analysis of studies in experimental mammals on litter size and fetal weight, respectively. Regarding possible delayed effects of *in utero* exposure, RF-EMF probably does not affect offspring brain weight and may not decrease female offspring fertility; on the other hand, RF-EMF may have a detrimental impact on neurobehavioural functions, varying in magnitude for different endpoints, but these last findings are very uncertain.
Further research is needed on the effects at birth and delayed effects with sample sizes adequate for detecting a small effect. Future studies should use standardized endpoints for testing prenatal developmental toxicity and developmental neurotoxicity (OECD TG 414 and 426), improve the description of the exposure system design and exposure conditions, conduct appropriate dosimetry characterization, blind endpoint analysis and include several exposure levels to better enable the assessment of a dose-response relationship.
Protocol registration and publication: The protocol was published in Pacchierotti et al., 2021 and registered in PROSPERO CRD42021227746 (https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=227746).

1. Introduction

1.1. Rationale

Radiofrequency electromagnetic fields (RF-EMF) are now employed worldwide in a variety of technologies and exposure of human populations is widespread in both occupational settings and in everyday life. National and international human exposure limits to RF-EMF have been developed and are periodically revised (ICNIRP 2020). Nevertheless, due to the speed of change of emerging technologies and to the research and reviews with varying results, public concern about possible adverse health effects from RF-EMF has been raised.

Within an ongoing project aimed at assessing potential health effects of exposure to RF-EMF in the general and working population, the World Health Organization (WHO) International EMF Project identified six priority topics on which to focus: cancer, adverse reproductive outcomes, cognitive impairment, self-reported symptoms, oxidative stress, and heat-related effects (Verbeek et al., 2021). On these topics, the WHO has commissioned systematic reviews of observational and experimental studies to analyse and synthesize the available scientific evidence.

When considering “adverse reproductive outcomes”, effects on both male fertility and pregnancy are considered. We recently published a protocol for conducting systematic reviews on these topics based on animal studies and studies on human sperm exposed *in vitro* (Pacchierotti et al., 2021). A protocol was also published for a systematic review of human observational studies on the same topics (Kenny et al., 2022). The results of our systematic reviews of experimental studies are published in two separate papers, the present one focusing on adverse pregnancy outcomes and the second one focusing on male fertility (in preparation).

Embryonic and fetal development is a critical stage of life known to be extremely sensitive to environmental influences. Adverse effects range from developmental delay reflected as weight reduction at birth to congenital defects and miscarriage. Low birth weight occurs in 14.6% of births (Blencowe et al., 2019). Each year over 3 million children worldwide are born with a congenital anomaly with complex and multifactorial etiology. Approximately 5% of birth defects are estimated to be attributable to environmental factors (Baldacci et al., 2018). Moreover, subtler developmental alterations, especially on neurocognitive system, may manifest later in life.

Over the years, several experimental studies have been conducted with laboratory animal models, essentially rodents, to investigate RF-EMF effects during pregnancy, which confirmed knowledge about the detrimental effects of heating on embryonic/fetal development (Ziskin and Morrissey 2011). These studies have been previously reviewed (AGNIR 2012, Heynick and Merritt 2003, ICNIRP 2020, O’Connor 1999, SCENIHR 2015) but, until now, no systematic review was conducted according to internationally standardized protocols, including risk of bias assessment.

1.2. Objective

To overcome the limitations of the current assessment of the scientific evidence on the RF-EMF effects on pregnancy and birth outcomes, we carried out a systematic review of peer-reviewed literature on these topics according to the guidelines of the WHO (WHO 2014) and of the National Toxicology Program/Office of Health Assessment and Translation (NTP/OHAT) Handbook (NTP 2015a). In particular, we formulated a PECO (Population, Exposure, Comparator, Outcome) statement in which the Population of interest corresponded to experimental mammals exposed exclusively *in utero*, the Exposure of interest consisted of RF-EMF in the frequency range 100 kHz − 300 GHz, the Comparators were animals sham-exposed or exposed to the same temperature increase induced by RF-EMF exposure by direct heating, and the Outcomes

were reduction of fecundity, adverse effects on the offspring health at birth or delayed effects on the offspring health, each one assessed by multiple non-redundant endpoints (Fig. 1).

2. Methods

The systematic review was conducted according to the protocol published in Pacchierotti et al., 2021. Where there were deviations, this is mentioned in Section 4.5.2 Deviations from the Protocol.

2.1. Eligibility criteria

The eligibility criteria applied to select studies for inclusion in the systematic review were those published in the protocol. Studies in which the exposure level could only be inferred from assumed exposure conditions and not by a measurement or estimate were assessed together with all the other studies because it was difficult to set boundaries in a continuum of exposure dosimetry reporting quality. This is a slight deviation from the protocol (see Section 4.5.2). Studies were screened in relation to each of the elements of the PECO statement as reported in Table 1.

We considered only original, controlled experimental studies on non-human mammals exposed *in utero*, published in peer-reviewed journals. We excluded non-experimental studies (e.g., human epidemiologic or other observational studies).

We excluded papers reporting reviews, opinions, proceedings or meeting abstracts. We did not impose any year-of-publication or language restriction.

2.2. Information sources

Three publication databases were searched for eligible studies: NCBI PubMed (https://pubmed.ncbi.nlm.nih.gov/), Scopus (https://www.scopus.com/) and EMF Portal (https://www.emf-portal.org/), a database maintained by the RWTH Aachen University, Germany, specifically focused on EMF effects. The three databases were last consulted on September 8, 2022 (NCBI PubMed and Scopus) or on September 17, 2022 (EMF Portal).

2.3. Search strategy

We interrogated the NCBI PubMed and SCOPUS databases, without any limits on year or language, by search queries composed by English terms identifying the exposure, the outcome and the population. We combined these elements in the queries by the Boolean operators "AND/OR/NOT" as described in the Supplementary File 5 of Pacchierotti et al., 2021. Search terms were identified to retrieve all relevant peer-reviewed publications of RF-EMF effects on adverse pregnancy outcomes and congenital disorders, by reviewing PubMed Medical Subject Heading (MeSH) terms associated with relevant papers and testing these and other terms chosen by expert judgement through an iterative trial-and-error process. The removal of non-experimental and human studies was done manually rather than by the use of search filters because studies might have been incorrectly indexed in the databases. We searched the EMF Portal database selecting pre-defined domains for topics, frequency ranges and time span among the options, and combining appropriate key words chosen from those listed in the Glossary (Supplementary File 5 of Pacchierotti et al., 2021). The search outputs were then aligned to exclude duplicates and the resulting list was screened for eligibility criteria.

The search strategy was peer-reviewed as part of the publication process of the protocol.

2.4. Selection process

Two reviewers independently evaluated the titles and abstracts of the identified papers to exclude records that were not relevant or did not fulfil one or more of the inclusion criteria for the PECO elements. In the case of disagreement between the reviewers, or when the abstract did not report enough information, we passed the paper to the full text evaluation phase.

Two reviewers independently evaluated the full texts of the

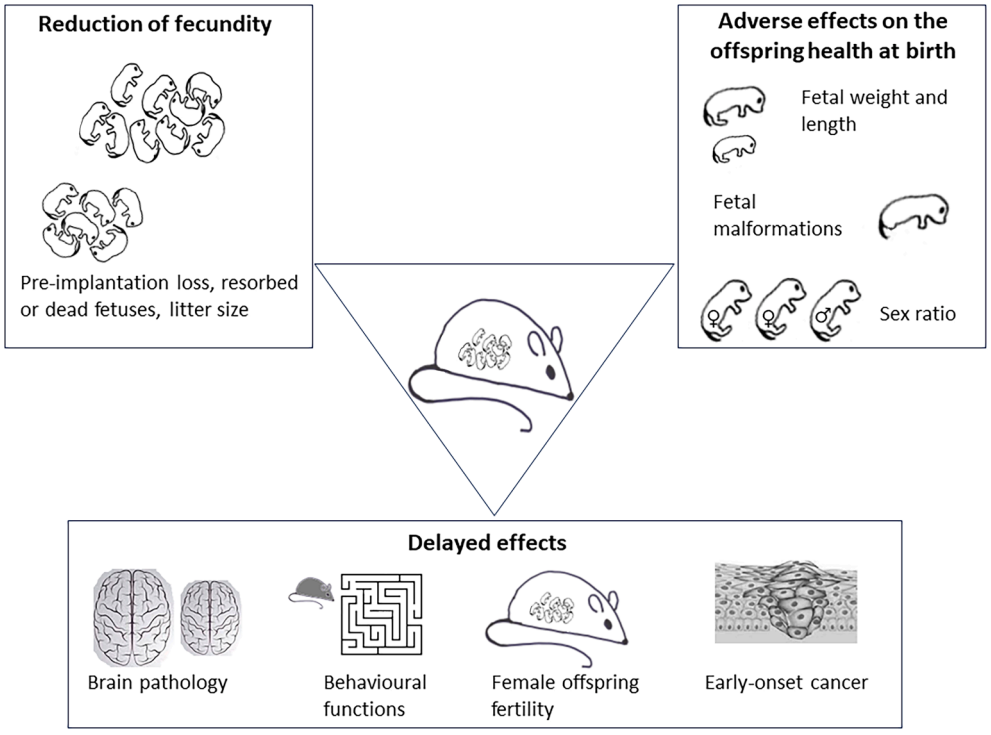

Fig. 1. Schematic presentation of the 3 outcomes considered in the systematic review: reduction of fecundity, adverse effects on the offspring health at birth and delayed effects on the offspring health. Each of the 3 outcomes includes multiple endpoints measured by different tests.

Table 1
Eligibility criteria.

PECO	*Inclusion criteria*	*Exclusion criteria*
Population	- Experimental mammals exposed exclusively *in utero*	- Non-mammalian species - Experimental mammals not exposed *in utero*
Exposure	- RF-EMF (frequency range 100 kHz – 300 GHz) at any exposure level - Electromagnetic pulses (EMP)	- Static or extremely low-frequency magnetic and/or electric fields - Optical radiation - Ultrasound - Magnetic Resonance Imaging (MRI) - Mobile phone not in GSM mode, and not controlled by hardware or software, unless supported by measured or calculated metrics as specified in the protocol (Pacchierotti et al., 2021) - Experimentally controlled co-exposure to RF-EMF and other chemical or physical agents (typically aimed at testing combined effects) - Exposure signals with more than 10% of the total signal energy outside the considered frequency range 100 kHz – 300 GHz - Exposure levels for which a minimum contrast between exposed and comparator groups was not guaranteed, as detailed in the protocol
Comparator	- Sham-exposed controls - Temperature controls	- Historical controls (if a study included both historical and matched controls, the study was included considering only the latter as the comparator) - Animals not handled as the animals of the exposed groups with particular reference to possible restraint conditions, anesthesia and stressing manipulations (as specified in the protocol)
Outcomes	- Reduction of fecundity (pre-implantation losses; resorbed/dead fetuses; decrease of litter size) - Adverse effects on the offspring health at birth (decrease of fetal weight or length; increase of malformations; changes of sex ratio) - Delayed effects on the offspring health (brain weight and neuropathology; alteration of behavioural ontogeny landmarks; adverse effects on learning and memory, motor activity or motor and sensory functions; early-onset cancer; female fertility impairment)	- Outcomes measured by methods deemed invalid as insufficiently described, insufficiently validated or improperly applied (as specified in the protocol) - Qualitative evidence of alterations in fetuses, pups or adult progeny - Molecular alterations in the progeny that are not causally linked to human pathologies - Effects in adult progeny for which there is not a solid knowledge regarding their possible developmental origin - Delayed effects at cellular or tissue levels in organs other than brain or ovary - Delayed effects when part of exposure occurred after birth

identified papers, and any disagreement between the reviewers was resolved by discussion or through involving a third reviewer. If findings from a study were described in more than one article, these were considered as one study only.

In no case was it necessary to contact the authors to decide about selection. Non-English language papers were either translated by the reviewers or through the use of Google Translate (https://translate.google.com/).

2.5. Data collection process

For all eligible studies, one reviewer extracted the study characteristics and results, and a second reviewer checked all the extracted information against the relevant article for completeness and accuracy as a quality control measure. If disagreement occurred between the reviewers, this was resolved through discussion or by consulting a third reviewer. In no case were the reviewers the authors of the scrutinized papers. When essential data were missing or there were inconsistencies in the reported information, the authors were contacted by e-mail and in case of no-reply, a reminder was sent. In cases where power density or other exposure metrics were reported instead of the whole body average SAR, the latter was estimated, if possible, on the basis of the available information. Data shown in figures were extracted by using digital rulers. In some cases, we re-calculated quantitative results from other data reported, to produce a form best suited to a *meta*-analysis, for example, converting standard errors into standard deviations or calculating means and variation parameters from raw data.

2.6. Data items (outcomes)

We extracted outcomes considered to be most representative of an effect on pregnancy and relevant for human health, following what was planned in the protocol. For the selection of outcomes we took into account the Organization for the Economic Co-operation and Development (OECD) Test Guidelines on prenatal developmental toxicity and developmental neurotoxicity studies (OECD TG 414 and 426).

We organized extracted data into three outcome categories, each one including multiple endpoints. The names for the outcomes and endpoints are slightly different from those used in the protocol because we wanted to use a terminology that better reflected the items actually reported in the literature, as detailed in Supplementary File 1.

Reduction of fecundity. This category included: measurements of pre-implantation losses, expressed as percentage or number of losses or as number of implants recorded right after the end of the implantation process; decrease of litter size or increased incidence of resorbed or dead fetuses measured just before or within 3 days after birth.

Adverse effects on the offspring health at birth. This category included reduction of fetal/pup weight or fetal/pup length, increase of external, visceral and skeletal malformations, ano-genital distance, change of sex ratio. Data on sex ratio were treated as a binary variable extracting odds ratio for the number of males as the effect size for the *meta*-analysis.

Delayed effects on the offspring health. This category included measurements of effects that became evident later in life after *in utero* exposure, in particular neurobehavioural alterations, effects on the female reproductive system and early-onset cancer. Among the possible markers of neurobehavioural alterations, data were extracted on learning and memory, motor activity or motor and sensory ability tests, behavioural ontogeny, brain weight and central nervous system histopathology. The age of first surface or air righting was considered the most representative biomarker of behavioural ontogeny alterations and synthesized in a *meta*-analysis. Reduction of the oocyte pool or of the litter size was considered evidence of effects on the female reproductive system, using the latter data for the *meta*-analysis.

While specific indications on which data to extract were provided in the protocol (Supplementary File 2 of Pacchierotti et al., 2021), in some cases it had not been possible to foresee the type of data reported in the papers and additional decisions, listed in Supplementary File 2, were taken before inspecting the results.

For the synthesis of results, primary outcomes considered were those measured by endpoints assessed at birth because of the social burden of perinatal deaths and congenital malformations in humans, and because the connection between prenatal exposure and adverse birth effects is closer than for delayed effects.

2.7. Data items (other variables)

In addition to outcome data, we also extracted information relating to the populations of experimental animals and the exposure conditions. In particular, the species, strain, age, number of animals (dams and offspring) were recorded. Age of dams at exposure and of pups at analysis were considered to assess the comparability of animal characteristics between comparator and exposed groups. Several variables were extracted to characterize exposure conditions and to assess the risk of bias: frequency, modulation, exposure system, exposure level, exposure duration, animal temperature, and period of gestation during which the exposure occurred.

Particular attention was given to the extraction of dosimetric information that defined the exposure level. When data on the exposure level (s) in terms of whole body average SAR were not reported, a SAR estimate was calculated based on other dosimetric information and biophysical assumptions. Randomization of animals to study groups, allocation concealment and blinding during exposure and/or outcome assessment, sham exposure conditions and statistical methods applied were considered as elements in the risk of bias assessment. We extracted but did not further analyse information on conflict of interest and funding sources, as initially planned in the protocol, since, in the vast majority of papers, public funding and absence of conflict of interest were declared (see Section 4.5.2).

2.8. Study risk of bias assessment

Risk of bias (RoB) was evaluated using the RoB Rating Tool developed by OHAT (NTP 2015a, b), with minor modifications informed by RoB expertise developed within SYRCLE (Hooijmans et al., 2014). Six bias domains were considered: 1) Selection bias; 2) Performance bias; 3) Detection bias relative to confidence in the exposure and outcome assessment; 4) Attrition/Exclusion bias; 5) Selective reporting bias; 6) Other sources of bias. For each of these domains a set of predefined questions guided the reviewers in the assessment of the internal quality of data. Supplementary File 3 summarizes the questions within each domain. Questions were based on those proposed in the OHAT handbook (NTP 2015a, b); the question "Has possible RF-EMF induced temperature increase been adequately considered and assessed?" was added because this aspect is especially relevant in the case of RF-EMF exposure to assess confidence in the exposure conditions. A customized guide to RoB assessment in the frame of the specific systematic review topic was developed to assist the reviewers as reported in the Supplementary File 10 of Pacchierotti et al., 2021.

Following the 3-tier system of study classification proposed by the OHAT, the scores for the different questions were integrated to obtain the study overall RoB estimate. A study was labelled "high concern" when one or more questions were answered with "definitely high RoB". A study was labelled "low concern" when none of the questions were answered with "probably high RoB" or "definitely high RoB". All other studies were labelled as "some concern".

Two reviewers independently analysed the included papers for RoB assessment, and disagreements were resolved by discussion with a third reviewer. RoB was evaluated at the endpoint level, meaning that one paper that reported results for different endpoints received multiple RoB evaluations. Whenever necessary to clarify issues relevant for RoB assessment, authors were contacted and their reply or absence of reply was considered in the assigned scores.

2.9. Effect measures

Original results were expressed as continuous or binary variables. When the same results were expressed in both ways, we extracted the continuous variable. We also tried to transform original binary data into continuous variables, but often this was not possible because of a lack of information. When possible, we transformed original data into the most common metrics but, in many cases, this was not possible because of lack of data.

For continuous variables, a Mean Difference (MD) was calculated. Standardized Mean Differences (SMD), calculated as MD/pooled SD, were used for data that used different metrics (percentages or numbers) to measure the same endpoint, or when the scale of measures was expected to widely differ, e.g., in the case of mouse or rat weight and length. For binary variables, we calculated Odds Ratios (OR) instead of Risk Ratios as initially planned in the protocol (see Section 4.5.2).

Whenever results were not clearly reported as litter means and considering that the experimental unit in adverse pregnancy studies is the individual dam, we adjusted them for litter clustering by applying an intracluster correction (ICC) factor of 0.2 in the design effect formula provided by Golub and Sobin (2020). Unfortunately, in many studies reporting delayed effects in the offspring, the number of dams was unavailable and we preferred not to introduce further assumptions and did not adjust data for litter clustering. This could have introduced a bias towards the detection of an RF-EMF effect, however, in the case of delayed effects, litter clustering is expected to have a small influence on the results and the bias should only have a minor impact on data synthesis.

2.10. Synthesis methods

All included papers were organized in a tabular form by the first author surname in alphabetical order. Several papers compared the effect of different exposures or exposure levels also to different sham exposed groups. Each of the exposure-sham combinations was considered as a separate study. Studies that were homogenous with regard to the PECO elements were synthesized. In particular, we made syntheses of results for 14 different endpoints belonging to the 3 outcomes as shown in Table 2.

For each endpoint we first conducted a *meta*-analysis of exposed vs sham control comparisons. When a study had several exposure groups matched to the same comparator, the means and standard deviations of these exposed groups were combined into one exposed group using the formulas provided in the paragraph 6.5.2.10 of the Cochrane Handbook (Higgins et al., 2022), so that each study was entered only once into the *meta*-analysis. The exposure level assigned to that combined exposed group was calculated as the average SAR of the exposed groups in that study weighed by the number of animals in each exposed group. In the forest plots this is indicated with an asterisk after the study ID. Studies that compared each exposed group to another separate sham control group were entered as separate studies in the *meta*-analysis. When multiple studies were reported in the same paper, this is indicated with a number after the study ID in the forest plot.

A random-effects *meta*-analysis model was used because the underlying effect size was expected to differ between studies due to the explorative nature and diversity of animal studies. Statistical heterogeneity of results was assessed by measures of heterogeneity variance (τ^2, I^2). For the random-effects model, the DerSimonian and Laird between-study variance estimator was used.

A forest plot was drawn in which the studies were divided according to their overall RoB level as "low or some concern" or "high concern". In order to assess possible RF-EMF impact upon the most robust data, we conducted all further analysis and assessed the certainty of evidence only for studies rated at "low or some concern" (see Section 4.5.2).

To explore possible causes of heterogeneity, we conducted sub-group analyses according to animal species, exposure levels (SAR less than 0.1, 0.1 $\leq$ SAR less than 5, SAR $\geq$ 5 W/kg) and measurements of animal core temperature increase below or above 1 °C. We limited the subgroup analysis to these 3 variables, because they were considered the most likely to affect a possible association between exposure and outcomes and to keep the work manageable. This is a slight deviation from the protocol (see Section 4.5.2).

According to the protocol, subgroup analyses were only interpreted

Table 2
List of endpoints for which the results have been synthesized by separate *meta*-analyses or narrative syntheses.

Endpoint	Metrics	Studies entered into a *meta*-analysis		Papers presented by a narrative synthesis
		N° papers (N° studies)[1]	Effect size measure	N° papers
Reduction of fecundity				
Pre-implantation loss	Pregnancy rate			1
	Mean number of implantation sites	1 (2)	SMD	1
	Mean losses per litter (%)	1 (2)		1
	Mean number of losses per litter	1 (3)		
Litter size	Mean number of fetuses/pups per litter	29 (47)	MD	8
Resorbed or dead fetuses	Total number of resorbed or dead fetuses	16 (76)	OR after adjustment for intra-litter clustering, when necessary	2
Adverse effects on the offspring health at birth				
Fetal weight	Mean weight per litter (g)	44 (104)	SMD after adjustment for intra-litter clustering, when necessary	8
Fetal length	Mean length per litter (mm)	10 (32)	SMD after adjustment for intra-litter clustering, when necessary	1
Fetal malformations[2]	Mean number of malformations per litter	9 (13)	SMD	5
	Mean malformed fetuses per litter (%)	6 (36)		
Litters with malformed fetuses[2]	Number of litters with malformed fetuses	3 (30)	OR after adjustment for intra-litter clustering, when necessary	
	Number of fetuses with malformations	11 (50)		
Sex ratio	Total number of males	10 (27)	OR after adjustment for intra-litter clustering, when necessary	2
Delayed effects on the offspring health				
Brain pathology	Mean brain weight (g)	8 (13)	SMD	
	Mean total number of Purkinje cells			1
	Mean number of Purkinje cells per mm[2]			2
	Mean hippocampal neuronal cell density (cells/optical field)			1
	Mean number of pyramidal neurons in hippocampus			2
	Mean spinal cord motor neuron number			1
Behavioural ontogeny	Mean post-natal day of first righting	4 (4)	MD	2
Learning and memory functions	Maze test - mean escape latency time (sec)	9 (14)	SMD	1
	Standard object recognition memory test - mean preference (%)			1
	Passive avoidance learning and memory test - mean time (sec)			3
	Conditioned avoidance response test - mean number of successful avoidances			1
	Food reinforcement learning test - mean efficiency (%)			1
	Maze test - mean number of total probe activities			1
	Lashley water maze test - mean number of errors			1
	Y maze test - mean number of shocks to reach performance			1
	Fear memory test - mean freezing time (%)			1
Motor activity functions	Mean time of endurance (sec)	7 (13)	SMD	
	Open field test - mean distance (cm)			4
	Open field test - mean number of crossed areas			5
	Open field test - mean crossed areas in last with respect to first trial (%)			1
	Locomotor activity - mean exploring time (%)			2
Motor and sensory functions	Mean startle magnitude (arbitrary units)	2 (4)	SMD	
Female infertility	Mean female offspring litter size	4 (5)	SMD	
	Mean number of follicles			2

Abbreviations: MD: Mean Difference; OR: Odds Ratio; SMD: Standardized Mean Difference.
[1] The number of studies here corresponds to the number of different exposure groups reported in the papers. This number may be higher than the number of studies analyzed in the results synthesis because when multiple exposure groups shared the same comparator, their data were averaged and considered as one study only.
[2] The term "malformed fetuses" includes fetuses carrying any type of major external, skeletal or visceral malformation.

when all the subgroups included at least 3 studies.

Next, we conducted a dose–response *meta*-analysis as described by Orsini and Spiegelhalter (2021) and implemented in STATA (STATA/BE 17.0 by StataCorp LLC, College Station, Tx USA, 2022). We specified a model based on an assumed linear relation between the whole body average SAR and the outcome. We also specified a non-linear model based on cubic splines. To assess if the non-linear model fit better than the linear model, we used the difference between the Akaike's information criterion (AIC) of the models. Finally, we visualised the summary estimate of the linear and the non-linear model together with the individual study dose–response curves in one graph based on the best linear unbiased prediction. We compared the predicted effects at a dose of 1 W/kg to the other doses over a range of 0 to 10 W/kg.

We used STATA 17 for the *meta*-analysis and the dose–response *meta*-analysis.

2.11. Reporting bias assessment

We assessed reporting publication bias in all the studies retrieved, irrespective of their overall RoB level of concern, to enlarge as much as possible the database and increase the sensitivity of our analysis. To visualise possible publication bias, funnel plots of the study effect size

measures against their standard errors were produced when at least 5 studies were available. If the funnel plot, upon visual inspection, showed that more imprecise studies with non-harmful effects were missing, this was considered an indication of possible publication bias. If ten or more studies were included in the same *meta*-analysis, an Egger's test was applied to evaluate potential small study bias, otherwise a qualitative evaluation was made (Egger et al., 1997).

2.12. Certainty assessment

The Grading of Recommendations, Assessment, Development and Evaluations (GRADE) framework for developing and presenting summaries of evidence was used to judge the certainty in the evidence of the effects observed in the systematic review and to draw conclusions (https://www.gradeworkinggroup.org). GRADE was initially developed for clinical studies and its application to animal toxicological studies is still under development. Toxicological studies pose a challenge to the GRADE approach because they are much less standardised than clinical studies. We started the rating of the certainty of the evidence at high certainty as is performed in human experimental studies (Hooijmans et al., 2018). We considered five domains: limitations in studies, indirectness considering how well the PECO question has been addressed from both the animal and human perspective, inconsistency, imprecision and publication bias. Depending on which criteria for which domains were met, we downgraded the certainty of the evidence to moderate, low, or very low according to the Supplementary File 11 of Pacchierotti et al., 2021. The only upgrading factor considered was consistency among animal species. Although we explored dose–response relationships, we did not apply evidence of a dose dependent effect as a further upgrading factor, because assessment of dose dependency was not considered by the PECO question and GRADE evidence profiling already started from high certainty, as indicated for experimental animal studies.

In analogy to the rating of importance of outcomes as described by Guyatt et al. (2011), we added a column to the GRADE evidence profile to report the level of importance we attributed to each endpoint in relation to the ultimate human relevant outcome. We used a scale from 1 (the lowest importance) to 10 (the highest importance) to assign a score to each endpoint. We used these scores for the assessment of the indirectness domain. For example, we attributed less importance to an alteration of the expected 1:1 sex ratio and more importance to embryonic/fetal death leading to decreased litter size in rodents and a higher probability of spontaneous abortions in humans, as the latter has a clearly more severe impact on human health than the former. Similarly, we attributed a slightly higher importance to an irreversible congenital malformation than to a symptom of developmental delay, such as reduced fetal weight of fetal length.

3. Results

3.1. Study selection

Fig. 2 shows the flow diagram from the initially retrieved references to the finally included papers, as per the PRISMA 2020 template (Page et al., 2021). After exclusion of duplicate records and of papers deemed not eligible based on title/abstract, a total of 236 papers remained for full-text assessment; we could not retrieve 11 papers and were unable to translate 10 papers. Of the remaining 215 papers, we excluded 127 after reading the full text. Therefore, the systematic review is based on a total of 88 papers.

3.2. Excluded studies

A list of the excluded papers is reported in Supplementary File 4 with a justification of the exclusion rationale. Most studies were excluded because of invalid or out-of-scope outcome assessment data. Examples were studies assessing effects in post-weaning offspring in organs unrelated to neurobehavioural effects, studies assessing effects on placenta

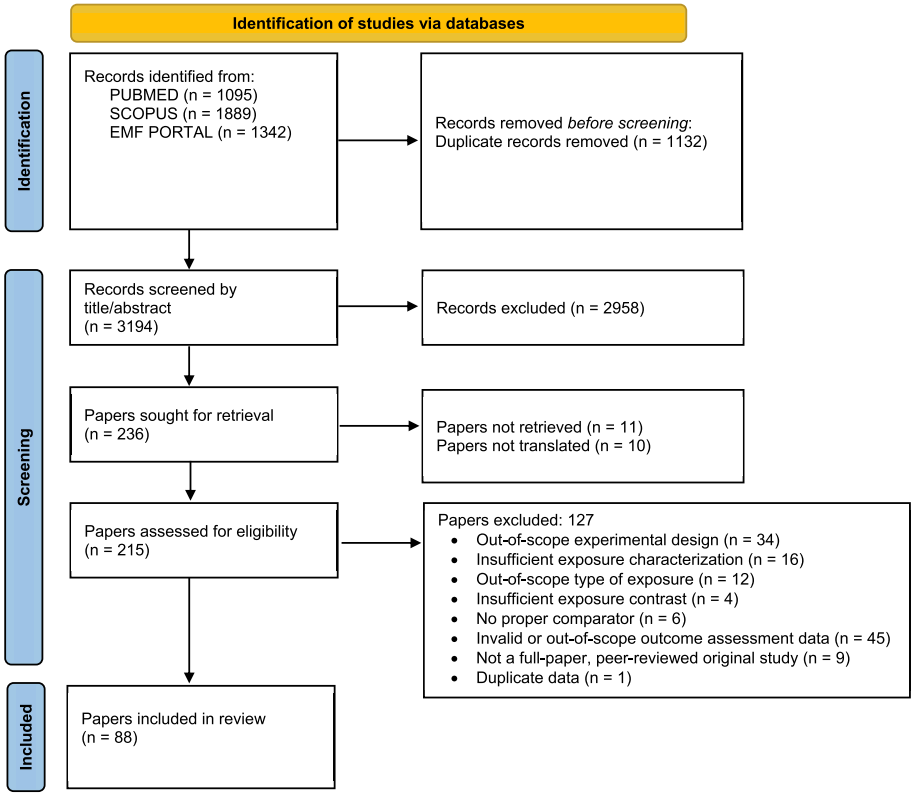

Fig. 2. Flow chart of the paper selection process according to the template proposed by PRISMA 2020.

only, or studies applying methods deemed invalid. Methods were considered invalid if insufficiently described, insufficiently validated or improperly applied (e.g., generic description of results without statistical support or immunohistochemical detection of cellular and molecular markers of stress). Several other studies were excluded because of out-of-scope experimental design, meaning that exposure was not or not only during pregnancy, that animals were co-exposed to RF-EMF and other beneficial or detrimental agents, or that animals were environmentally exposed under uncontrolled conditions. The same file also includes the references of the studies that could not be retrieved or translated, which had not been foreseen at the protocol stage.

3.3. Study characteristics

All papers that met the inclusion criteria were reported in tabular form (Table 3). In this table, characteristics of the studies regarding populations, exposure and outcomes are presented. Additionally, a very brief description of the main results in scope for the systematic review is reported, based on the authors' interpretation and discussion. We also summarised temperature increase in the exposed animals, when determined. The table also shows the specific outcomes investigated and whether the data were entered into the *meta*-analyses. A few results could not be synthesized in the form of *meta*-analysis because actual data were not reported and could not be retrieved even after contacting the authors. In some other cases, we extracted the results, but we did not enter them into a *meta*-analysis because we preferred not to increase the heterogeneity among study results by pooling very different types of tests. For the *meta*-analysis we selected the best standardized and most often reported test for measuring each endpoint. When this choice had not been specified at the protocol stage, we took the decision before inspecting the study results as detailed in Supplementary File 2. Finally, the results reported by Cobb et al. (2000) were not entered into a *meta*-analysis because they were the only results on the effects of electromagnetic pulses. Nevertheless, all these results have been synthesized in a narrative way.

3.3.1. Population

Of the 88 included papers, 65 reported results in rats, 20 in mice and 3 papers reported studies in other species. The average size of the experimental dam groups ranged between 3 and 82 with a median of 10.

3.3.2. Exposure

Eighty-six papers reported studies that included the organogenesis phase in the exposure window and 40 of them exposed all 3 phases of pregnancy: pre-implantation, organogenesis and late gestation. Of the included papers, 81 tested frequencies below 6000 MHz, 6 tested frequencies of 6000 MHz or higher and 1 reported results obtained with electromagnetic pulses. Forty-five percent of studies in which the information was reported applied modulated signals. A quarter of the papers tested more than one exposure level; the range of whole body average SAR values was extremely wide, between 0.00011 and 115 W/kg. Similarly, the duration of exposure ranged from a few minutes for just 1 day at a high exposure level to 24 h per day for the duration of gestation. In about half of the papers, animal temperature was measured and reported, with half of these recording a dam core body temperature increase equal to or higher than 1 °C.

3.3.3. Comparators

All papers included a control group that was considered sufficiently sham-exposed to be used as a comparator for the RF-EMF exposed animals. No study included a control group of animals exposed only to a direct temperature increase strictly comparable to that induced by RF-EMF.

3.3.4. Outcomes

Reduction of fecundity. Six papers investigated the possibility of an increase of pre-implantation losses. In one paper, the effect was assessed by the total number of implants measured right after the end of the implantation period (Alchalabi et al., 2016). In another paper, the effect of pre-implantation exposure was assessed by the pregnancy rate (Nawrot et al., 1985). Other papers reported results on the number or percentage of losses in relation to the number of corpora lutea (Berman et al., 1992, Lary et al., 1982, Tofani et al., 1986) or simply stated no impact on implantation without details on the assessment method (Lee et al., 2009). Three of the 6 papers reported data that were suitable for use in *meta*-analysis. Results on litter size and incidence of resorbed/dead fetuses were reported in 37 and 18 papers, respectively.

Adverse effects on the offspring health at birth. Several different endpoints contributed to the evaluation of offspring health at birth. Indicators of a deterioration of offspring health were: a decrease of fetal/pup weight (52 papers) or length (11 papers); an increase of external or internal (visceral or skeletal) malformations (34 papers); a deviation from the expected sex ratio (12 papers). As can be seen from Table 3, most of the papers reported results on these endpoints in a form suitable for *meta*-analysis. No study was retrieved on ano-genital distance.

Delayed effects on the offspring health. None of the included papers addressed carcinogenic effects in the offspring exposed *in utero*. Thirty-three papers dealt with delayed effects, either on neurocognitive development or on the female reproductive system. Delayed effects were assessed as early as 7 days after birth up to over 100 days of age, with most studies concentrating on the period between 30 and 60 days of age. Eight papers reported measurements of brain weight. In addition, 7 papers contained data on cellular density in organs of the central nervous system. Six papers contained data on changes in the onset of developmental landmarks like age of surface righting, which is part of the labyrinthine righting reflex. In 21 papers, neurobehavioural functional impairment was assessed by a variety of tests. These were grouped into 3 main categories: tests measuring learning and memory capacities (18 papers), tests measuring motor activity (18 papers) and tests measuring motor and sensory functions (2 papers). Six papers reported data on the fertility of the female offspring, either by the number of ovarian follicles (2 papers) or by functional tests (4 papers). Possible RF-EMF impact on male fertility after prenatal exposure is reviewed in a separate paper.

Table 2 shows for each endpoint and outcome the number of papers and studies, the metrics in which the results were expressed, and the effect size measures used for the synthesis of results by *meta*-analysis.

3.4. Risk of bias in studies

Supplementary File 5(a-n) shows, for each endpoint, the heatmaps of the consensus scores assigned to each RoB question together with the overall level of concern; the relative justifications are reported in Supplementary File 6. In these Supplementary Files each entry within a given endpoint might include multiple comparisons between sham and exposed groups.

3.4.1. Reduction of fecundity

3.4.1.1. Pre-implantation loss. Four studies were classified at "some concern", 2 studies were classified at "high concern". The main reasons for concern were limited confidence in the outcome assessment and lack of blinding during experiment performance, and, for the "high concern" studies, insufficient exposure characterization and/or inadequate assessment of temperature issues.

3.4.1.2. Litter size. Twenty-four studies were classified at "some concern", 11 studies were classified at "high concern" and 3 studies were classified at "low concern". The main reasons for "some concern" were limited confidence in the outcome assessment and lack of blinding during experiment performance. The main reasons for "high concern"

Table 3
List of included papers with main study characteristics.

Reference	Population		Exposure			Outcome			
	Species (Average group size)	*Stage of prenatal development during exposure: pre-implantation (PI), organogenesis (O), late gestation (LG)*	*Frequency (MHz)/ Modulation (M, CW) or EMP*	*Average level(s) W/kg*	*Duration(s) Hours per day /N° of days*	*Fecundity*	*Health at birth*	*Delayed effects*	*Summary of paper results*
Aït-Aïssa et al., 2012	Rat (10)	O + LG	2450/M	0.08, 0.4, 4	2:00/11	X			No effect on litter size and pup weight.
Albert et al., 1981	Rat (3/3)	LG	2450/CW	2	21:00/5				Decrease of Purkinje cell number 40 days after birth. Pups as the experimental unit.
Alchalabi et al., 2016	Rat (10)	PI, PI + O + LG	1800/M	0.048	1:00, 2:00/7, 20	X	X		Decrease of implantation sites after the longest daily exposure during preimplantation stage. Decrease of litter size and fetal weight after the whole gestation exposure.
Alchalabi et al., 2017	Rat (20)	PI + O + LG	1800/M	0.974	1:00, 2:00/20				Variable effects on skeletal development and malformations.
Aldad et al., 2012	Rat (37/77)	PI + O	800–1900	1.6	24:00/17				At PND 56–112 impairment of memory by standard object recognition test. Effect observed also in tests measuring hyperactivity and anxiety but not in test measuring fear. Pups as the experimental unit.
Anderson et al., 2004	Rat (25)	LG	1620/M	0.06	2:00/3				No effect on litter size.
Azimzadeh and Jelodar, 2020	Rat (7/NR)	PI + O + LG	900/CW	0.035	4:00/21				At PND 45 effect on learning and memory by passive avoidance test.
Bas et al., 2013	Rat (3/6)	O + LG	900/CW	0.01*	1:00/9			X	At PND 32 no effect on brain weight; decrease of hippocampus pyramidal cell number. Pups as the experimental unit.
Berman et al., 1978	Mouse (82)	PI + O, O	2450/CW	2, 7, 8.1, 22.2*	1:40/10, 17	X	X		0.1 °C temperature increase at the highest exposure level, no increase at the other exposure levels. No effect on litter size; decrease of fetal weight at the highest exposure level; increase of % litters with external malformations at the lowest exposure level.
Berman et al., 1981	Rat (65)	O	2450/CW	4.2	1:40/10	X	X		2 °C temperature increase. No effect on litter size and on fetal weight; no effect on external, visceral or skeletal anomalies or variations.
Berman et al., 1982a	Mouse (15)	O	2450/CW	16.5	1:40/12	X	X		0.1 °C temperature increase. No effect on litter size; decrease of fetal weight; delay of sternal ossification.
Berman et al., 1982b	Hamster (44)	O	2450/CW	6, 9	1:40/9	X	X		0.4 and 1.6 °C temperature increase at the 2 exposure levels. No effect on litter size; decrease of fetal weight, increase of fetal resorptions and delay of sternal ossification at the highest exposure level. No effect on internal or external anomalies.
Berman et al., 1984a	Rat (21)	O	2450/CW	6	1:40/10	X	X		2 °C temperature increase. No effect on litter size; decrease of fetal weight; delay of sternal ossification.
Berman et al., 1984b	Mouse (20, 10/10)	O	2450/CW	16.5	1:40/12		X	X	0.1 °C temperature increase. Decrease of pup weight. At PND 10–17 decrease of brain weight.
Berman et al., 1992	Rat (32)	PI + O + LG	970/CW	0.07, 2.4, 4.8	22:00/19	X	X		Temperature not measured, but highest exposure level considered likely hyperthermal. No effect on preimplantation loss. No effect on litter size; decrease of fetal weight at the highest exposure level; no effect on sternal ossification.
Bornhausen and Scheingraber, 2000	Rat (12, 10/10)	PI + O + LG	900/M	0.046	24:00/20				No effect on litter size. At PND 80 no evidence of cognitive impairment by food reinforcement learning test, pups as the experimental unit.
Brown-Woodman and Hadley, 1988a	Rat (4)	O	27.12/CW	11.2	0:01, 0:02, 0:03, 0:05, 0:10, 0:15, 0:18, 0:20, 0:25, 0:30, 0:50, 1:00/1	X	X		2.5–5 °C temperature increase in the various exposed groups. Effects on fetal weight, embryolethality and malformations, increasing with the temperature elevation and increasing exposure time.
Brown-Woodman and Hadley, 1988b	Rat (7)	O	27.12/M	2.8, 3, 4.2, 4.3, 5.2, 5.6*	0:30, 0:45, 1:00/1	X	X		0.4–1.3 °C temperature increase in the various exposed groups. Embryolethal effects depending on the exposure duration and pulse repetition frequency; no effect on fetal weight and fetal malformations.
Calis et al., 2019	Rat (NR/4)	PI + O + LG	2100	0.23	1:00/21				At PND 42 in female offspring decrease of primordial and secondary follicle numbers.

(*continued on next page*)

Table 3 (*continued*)

Reference	Population		Exposure			Outcome			
	Species (Average group size)	*Stage of prenatal development during exposure: pre-implantation (PI), organogenesis (O), late gestation (LG)*	*Frequency (MHz)/ Modulation (M, CW) or EMP*	*Average level(s) W/kg*	*Duration(s) Hours per day /N° of days*	*Fecundity*	*Health at birth*	*Delayed effects*	*Summary of paper results*
Chazan et al., 1983	Mouse (38)	PI + O + LG	2450/CW	0.5, 2*	2:00/18	X			Increased embryolethality at the highest esposure level.
Chernovetz et al., 1975	Mouse (5, 12/12)	O	2450/M	38	0:10/1	X	X		2 °C temperature increase. No effect on litter size and malformations. At PND 38 no effect on learning by Lashley water maze test, pups as the experimental unit.
Chernovetz et al., 1977	Rat (15)	O	2450/M	31	0:20/1	X	X		3.4 °C temperature increase. Decrease of fetal weight; increase of resorptions with respect to both sham controls and infrared temperature controls.
Chiang 1988	Mouse (10, 10/21)	PI + O + LG	3000/M	3.25	5:00/20			X	No temperature increase. No effect on pup weight at PND 3. At PND 21 no effect on surface righting, learning and memory by maze test and motor activity by forelimb hanging test, pups as the experimental unit.
Cobb et al., 2000	Rat (6, 6/36)	PI + O + LG	EMP	0.045	0:02/16				No effect on litter size¸ fetal weight and sex ratio. No effect on air righting, locomotor activity and spatial learning by water maze test.
Dasdag et al., 2000	Rat (12)	PI + O + LG	890–915/M	0.155	0:03/20		X		0.317 °C temperature increase. No effect on litter size. Decrease of fetal weight with pup as the experimental unit.
DastAmooz et al., 2018	Rat (NR/6)	PI + O + LG	2450/M	0.23	6:00/24			X	At PND 56 effect on learning but not on memory by Morris water maze test; no effect on locomotor activity by open field test; pup as the experimental unit.
Erdem Koç et al., 2016	Rat (3/6)	PI + O + LG	900/M	2	1:00/21				At PND 28 decrease of hippocampus pyramidal cell number, pup as the experimental unit.
Ferreira et al., 2006	Rat (5)	PI + O + LG	834/M	0.89	8:30/21	X			No effect on litter size.
Galvin et al., 1983	Rat (20)	PI + O + LG	2450/CW	2.7	3:00/17		X		No temperature increase. No effect on pup weight, pup as the experimental unit.
Galvin et al., 1986	Rat (10, NR/ 15)	PI + O + LG	2450/CW	2	3:00/16	X	X	X	No effect on litter size and on pup weight with pup as the experimental unit. At PND 30 effect on swim endurance in males and in females in 1 out of 2 experiments; by the same test, no effect at PND 100; effects on motor and sensory outcomes by airpuff startle response in female but not in males; pup as the experimental unit.
Guler et al., 2010	Rabbit (9)	O + LG	1800/M	0.008*	0:15/8		X		No effect on pup weight.
Haghani et al., 2013	Rat (10, 10/10)	PI + O + LG	900/M	0.7	6:00/21	X		X	No temperature increase. No effect on litter size. On PND 30–32 no effect on motor activity by rotarod performance, wire grip and open field tests; pup as the experimental unit.
Ikinci et al., 2013	Rat (3/12)	O + LG	900	0.4*	1:00/9			X	At PND 30 effect on learning and memory by radial arm maze test and passive avoidance test; pup as the experimental unit.
Inaloz et al., 1997	Rat (8)	PI + O + LG	2450/CW	1.9, 3.9	0:15, 0:30/21	X	X		1 °C temperature increase. No effects on litter size, fetal weight and fetal length.
Inouye et al., 1983	Rat (6)	PI + O + LG	2450/CW	1.76	3:00/18		X		No effect on pup weight and incidence of malformations; no major effect on sex ratio.
Jensh et al., 1982a	Rat (7, 7/36)	PI + O + LG	915	3.5*	8:00/14	X	X	X	No temperature increase. No effect on litter size; increase of pup weight; no effect on malformations. Anticipation of surface righting; at PND 60–90 no effect on learning by water maze test; no effect on motor activity by open field and forelimb hanging tests; no effecto on brain weight; no effect on female fertility by F2 litter size; pup as the experimental unit.
Jensh et al., 1982b	Rat (7)	PI + O + LG	915	3.5*	8:00/14	X	X		No temperature increase. No effect on litter size; no effect on fetal weight and malformations.
Jensh et al., 1983a	Rat (7)	PI + O + LG	2450/CW	4.4*	8:00/14	X	X		No temperature increase. No effect on litter size; decrease of fetal weight; no effect on malformations.

(*continued on next page*)

Table 3 (*continued*)

Reference	Population		Exposure			Outcome			
	Species (Average group size)	*Stage of prenatal development during exposure: pre-implantation (PI), organogenesis (O), late gestation (LG)*	*Frequency (MHz)/ Modulation (M, CW) or EMP*	*Average level(s) W/kg*	*Duration(s) Hours per day /N° of days*	*Fecundity*	*Health at birth*	*Delayed effects*	*Summary of paper results*
Jensh et al., 1983b	Rat (8, 8/35)	PI + O + LG	2450/CW	4.4*	8:00/14		X	X	No temperature increase. Increase of pup weight; no effect on malformations. No effect on surface righting. At PND 60–90 no effect on learning and memory by shuttle box test; no effect on learning by water maze test; no effect on motor activity by open field and forelimb hanging tests; no effect on brain weight; increase of female fertility by F2 litter size; pup as the experimental unit.
Jensh 1984a	Rat (10, 8/29)	PI + O + LG	6000	7.28	8:00/14	X	X	X	No temperature increase. No effect on litter size; decrease of pup weight; no effect on malformations. No effect on surface righting; at PND 60 no effect on learning by water maze test; no effect on motor activity by open field and forelimb hanging tests; no effect on brain weight; no effect on female fertility by F2 litter size; pup as the experimental unit.
Jensh 1984b	Rat (9)	PI + O + LG	6000	7.28	8:00/14	X	X		No temperature increase. No effect on litter size and malformations; decrease of fetal weight.
Kaplan et al., 1982	Squirrel monkey (10, 6/4)	O + LG	2450/M	0.034, 0.34, 3.4	3:00/35–95				No effect on litter size. No effect on locomotor activity at 6–8 weeks of age.
Keles and Sut, 2021	Rat (NR/6)	O + LG	900	0.01*	1:00/9				At PND 32 decrease of spinal cord motor neuron number; pup as the experimental unit.
Kubinyi et al., 1996	Mouse (21, NR/ 236)	PI + O + LG	2450/M	4.23	1:40/19		X	X	No effect on pup weight. At PND 24 no effect on brain weight; pup as the experimental unit.
Lary et al., 1982	Rat (22)	PI, O	27.12/CW	11.1–12.5	0:26–0:32/1	X	X		4.4 °C temperature increase. No effect on pre-implantation loss. Increase of dead or resorbed fetuses after exposure of gestation day 7 or 9. Decrease of fetal weight after exposure of organogenesis stage but not after exposure of pre-implantation stage; decrease of fetal length; increase of external, visceral and skeletal malformations mainly after exposure of organogenesis stage. No effect on sex ratio.
Lary et al., 1983a	Rat (33)	O	100/CW	0.41	6:40/6		X		No temperature increase. No effect on fetal weight or length; no effect on skeletal malformations; no effect on sex ratio.
Lary et al., 1983b	Rat (23)	O	27.12/CW	10.8	0:18, 0:23, 0:38, 2:18/1	X	X		2.9–3.9 °C temperature increase as a function of exposure duration. Teratogenic and embryotoxic effects increasing as a function of temperature increase and exposure duration.
Lary et al., 1986	Rat (21)	O	27.12/CW	10.8	0:10, 0:17, 0:24, 0:31, 0:40/1	X	X		2.6, 3, 3.6, 4 or 4.5 °C temperature increase. Increase of embryotoxicity and teratogenicity starting at 3.6 °C temperature increase. No effect on sex ratio.
Lee et al., 2009	Mouse (17)	PI + O	848.5, 848.5 + 1950/M	2, 4	1:30/17	X	X		No temperature increase. No effect on implantation. No effect on litter size, fetal weight, fetal length, malformations and sex ratio.
Li et al., 2020	Rat (4, 4/18)	PI + O + LG	1800, 2400, 1800 + 2400	0.02, 0.2, 0.4*	12:00/21		X	X	Some effects on pup weight but no on pup length. At PND 21 no major effect on learning and memory by Y-maze test; at PND 49 no major effect on motor activity by open field test; pups as the experimental unit.
Marcickiewicz et al., 1986	Mouse (40)	PI + O + LG	2450/CW	4.5, 17*	2:00/18	X	X		1.5–2 °C temperature increase at the highest exposure level. Increase of embryotoxicity at the highest exposure level. Decrease of fetal weight, no effect on malformations; fetuses as the experimental units.
Merritt et al., 1984	Rat (10)	PI + O + LG	2450/M	0.4	24:00/17	X	X		No effect on litter size and fetal weight.
Nawrot et al., 1981	Mouse (20)	PI + O, O	2450/CW	6.7, 28.14, 40.2	8:00/6, 10, 15	X	X		1 and 2.5 °C temperature increase at the 2 higher exposure levels. Decrease of implantation sites per litter and fetal weight after exposure of the pre-implantation phase at the highest level; increase of malformations after exposure during the organogenesis phase at the highest level. Fetuses as the

(*continued on next page*)

Table 3 (*continued*)

Reference	Population		Exposure			Outcome			
	Species (Average group size)	*Stage of prenatal development during exposure: pre-implantation (PI), organogenesis (O), late gestation (LG)*	*Frequency (MHz)/ Modulation (M, CW) or EMP*	*Average level(s) W/kg*	*Duration(s) Hours per day /N° of days*	*Fecundity*	*Health at birth*	*Delayed effects*	*Summary of paper results*
									experimental unit. The study also includes a direct heating temperature comparator.
Nawrot et al., 1985	Mouse (22)	PI + O, O	2450/CW	40.2	8:00/6, 10	X	X		2.3 °C temperature increase. Decrease of pregnancy rate and decrease of fetal weight after exposure during pre-implantation phase; no increase of embryolethality in pregnant dams; no effect on external, visceral and skeletal malformations. Fetuses as the experimental unit. The study also includes a direct heating temperature comparator.
Nelson et al., 1991	Rat (22)	O	10/CW	Variable	0:30/1	X	X		One exposure level at 3.7 °C temperature increase achieved varying SAR between 0.8 and 6.6 W/kg. No effect on litter size and on fetal weight; increase of malformations; no effect on sex ratio.
Nelson et al., 1994	Rat (6)	O	10/CW	Variable	0:30, 0:40, 0:50, 1:00/1				Exposure level at 4 °C temperature increase achieved varying SAR between 0.8 and 6.6 W/kg. Increased incidence of resorptions more evident after exposure of gestation day 9 than of gestation day 13; decrease of fetal weight; effects on external, skeletal and visceral malformations.
Nelson et al., 1997a	Rat (10)	O	10/CW	Variable	Variable/1				Three exposure levels at temperature increase of 0.5 °C (for 0:20, 2:20, 4:20 or 6:20), 1.5 °C (for 0:20, 1:20, 2:20, 3:20 or 4:20) or 2.5 °C (for 0:20 or 1:20) achieved by varying SAR between 0.8 and 7.9 W/kg. Increase of malformations as a function of exposure duration at 1.5 °C temperature increase.
Nelson et al., 1997b	Rat (10)	O	10/CW	Variable	0:40/1		X		One exposure level at 4 °C temperature increase achieved varying SAR between 0.8 and 6.6 W/kg. Increase of external malformations.
Nelson et al., 1999	Rat (10)	O	10/CW	Variable	1:20/1	X	X		One exposure level at 3 °C temperature increase achieved varying SAR between 0.8 and 7.9 W/kg. Increased incidence of resorptions and malformations; no effect on fetal weight; fetuses as the experimental unit.
Nelson et al., 2001	Rat (10)	O	10/CW	Variable	1:20/1				One exposure level at 3 °C temperature increase achieved varying SAR between 0.8 and 7.9 W/kg. Increased incidence of resorptions and malformations; no effect on fetal weight; fetuses as the experimental unit.
O'Connor 1988	Rat (4, 4/NR)	PI + O + LG	2450/CW	270–310 W/m^2	6:00/19				2 °C temperature increase. Suggestion of a decrease of litter size. No effect on righting; at PND 36 no effect on motor activity by open field test.
Odaci et al., 2013	Rat (3/11)	O + LG	900	0.01*	1:00/9			X	At PND 26 effect on motor activity by rotarod test but not by open field test. Pups as the experimental unit.
Odaci et al., 2016	Rat (3/6)	O + LG	900/CW	0.01	1:00/9			X	At PND 32 no effect on cerebellum weight; decrease of Purkinje cell number. Pups as the experimental unit.
Ogawa et al., 2009	Rat (20)	O	1950/M	<0.2, 0.2	1:30/11	X	X		No effect on litter size, fetal weight, external, visceral and skeletal malformations, sex ratio.
Petitdant et al., 2018	Rat (8, 8/8)	PI + O + LG	900/M	0.7, 2.6	0:45/19			X	No effect on litter size and pup weight. At PND 36 no effect on learning and memory by freeze post-conditioning test; no effect on motor and sensory functions by acoustic startle test. Effect on motor activity by the open field test at the highest exposure level in adolescent but not in adult offspring. Pups as the experimental unit.
Poulletier de Gannes et al., 2012	Rat (14)	PI + O + LG	2450	0.08, 0.4, 4	2:00/17	X	X		No effect on litter size, malformations and sex ratio.
Razavinasab et al., 2016	Rat (NR/5)	PI + O + LG	900/M	0.6	6:00/21			X	No temperature increase. At PND 30 effect on learning and memory by water maze and passive avoidance tests. Pups as the experimental unit.
Rifat et al., 2016	Mouse (NR)	O + LG, PI + O + LG	10000	0.179	2:00/8, 20				No effect on litter size and pup length, decrease of pup weight.
Rugh and McManaway 1976	Mouse (28)	O	2450/CW	47, 85, 87*	0:04/1	X	X		1.5 °C temperature increase in the group exposed to 85 W/kg only. Embryotoxic and teratogenic effects lower in anesthetized than in non-anesthetized animals.

(*continued on next page*)

Table 3 (*continued*)

Reference	Population		Exposure			Outcome			
	Species (Average group size)	*Stage of prenatal development during exposure: pre-implantation (PI), organogenesis (O), late gestation (LG)*	*Frequency (MHz)/ Modulation (M, CW) or EMP*	*Average level(s) W/kg*	*Duration(s) Hours per day /N° of days*	*Fecundity*	*Health at birth*	*Delayed effects*	*Summary of paper results*
Rugh and McManaway 1977	Mouse (11)	PI, O	2450/CW	99.12–114.6	0:04/1	X	X		Effects on embryolethality and teratogenicity after exposure at some but not other days of gestation. Fetuses as the experimental unit.
Rugh and McManaway 1978	Mouse (NR/31)	O	2450/CW	106.33, 109.87	0:04/1			X	About 3 °C temperature increase. No effect on female offspring fertility. Pups as the experimental unit.
Sambucci et al., 2010	Mouse (11)	PI + O + LG	2462/M	4	2:00/14	X	X		Adjustment of TEM cell temperature to counteract animal temperature increase. No effect on litter size and pup weight.
Sangun et al., 2015	Rat (4)	PI + O + LG	2450/M	0.143	1:00/21		X		No effect on pup weight.
Schmidt et al., 1984	Rat (20)	PI + O + LG	2450/M	0.4	24:00/17	X	X		No effect on resorptions, fetal weight and fetal length. No effect on external and skeletal malformations.
Sharma et al., 2017	Mouse (6, 6/6)	O + LG, PI + O + LG	10000	0.179	2:00/9, 20	X	X	X	No effect on litter size and pup length; decrease of pup weight. At PND 21 decrease of brain weight after exposure during whole gestation; decreased number of Purkinje cells under all exposure conditions; pups as the experimental unit.
Shibkova et al., 2015	Mouse (16)	O	925/M	0.4*	0:10/5	X			No effect on litter size.
Shirai et al., 2014	Rat (8)	O + LG	2140/M	0.037, 0.11	20:00/16	X	X		No effect on litter size, malformations, sex ratio.
Shirai et al., 2017	Rat (12)	O + LG	Multiples, range: 870-5180/M	0.087, 0.433	20:00/16				No effect on litter size, malformations, sex ratio.
Smialowicz et al., 1981	Rat (20)	O + LG	100/CW	2.02	4:00/16		X		No effect on pup weight.
Smialowicz et al., 1982	Rat (6)	O + LG	425/CW	4.7	4:00/10		X		No temperature increase. No effect on pup weight.
Stasinopoulou et al., 2016	Rat (15, NR/9)	PI + O + LG	1900/M	0.018	12:00/21	X	X		No effect on litter size, pup weight, pup length. At PND 21 decreased hippocampal neuron density; pups as the experimental unit.
Takahashi et al., 2010	Rat (12)	O + LG	2140	0.029, 0.068	20:00/15	X	X		No effect on litter size, malformations and sex ratio.
Tofani et al., 1986	Rat (16)	PI + O, PI + O + LG	27.12	0.00011	24:00/7, 21	X	X		No temperature increase. No effect on pre-implantation loss rate. After whole gestation exposure increase of post implantation losses, no effect on external malformations, increase of visceral and skeletal malformations.
Türedi et al., 2016	Rat (3/6)	O + LG	900/CW	0.01	1:00/9				At PND 34 decrease of primordial follicle number in female offspring. Pups as the experimental unit.
Wang et al., 2018	Rat (6)	PI + O	850–1900/M	1.6	6:00, 24:00/17	X	X		No effect on litter size and on pup weight.
Wyde et al., 2018	Rat (5)	O	900/M	10, 12	9:10/6				About 1 and 2 °C temperature increase at 10 and 12 W/kg, respectively. No effect on the number of live fetuses and fetal weight; increased number of resorptions at the highest exposure level when modulated by GSM signal but not by CDMA signal.
Zhang et al., 2015	Mouse (NR/10)	PI + O + LG	9417	2	12:00/16			X	At PND 35 no effect on motor activity by open field test; effect on males but not on females on learning and memory by water maze test. Pups as the experimental unit.
Zhao et al., 2005	Mouse (24/144)	O	37400, 42200, 53000, 60000/ CW	0.45, 1.35, 2.25, 3.6	2:00/10				No temperature increase. At PND 60 effect on learning and memory by Y-maze test generally more evident at higher exposure levels. Pups as the experimental unit.

When data were entered into a *meta*-analysis this is marked by X. For studies on delayed effects, both average number of dams and average number of pups per experimental group have been reported as (N dams/N pups). When data were statistically analysed clearly considering fetuses/pups as the experimental unit this has been remarked. Information on modulation is not reported when it was not clearly provided in the paper. Exposure level(s) are expressed in SAR unless this data was not provided in the paper or calculated on the basis of other information, in which case PD values are reported if provided.

Abbreviations: CW: continuous wave unmodulated signal; EMP: electromagnetic pulses; LG: late gestation; M: modulated signal; NR: not reported; O: organogenesis; PI: pre-implantation; PND: post-natal day.

* SAR calculated for 1 or more of the exposed groups from different original exposure metrics.

were insufficient exposure characterization, inadequate assessment of temperature issues, and/or the small size of the experimental groups considered in the "other potential threats to internal validity".

3.4.1.3. Resorbed or dead fetuses. Fifteen studies were classified at "some concern" and 3 studies were classified at "high concern". The main reasons for "some concern" were limited confidence in the outcome assessment and lack of blinding during experiment performance. The reason for "high concern" was the small size of the experimental groups considered in the "other potential threats to internal validity".

3.4.2. Adverse effects on the offspring health at birth

3.4.2.1. Fetal weight. Thirty-eight studies were classified at "some concern", 13 studies were classified at "high concern" and 1 study was classified at "low concern". The main reasons for "some concern" were limited confidence in the outcome assessment and lack of blinding during experiment performance. The main reasons for "high concern" were insufficient exposure characterization, inadequate assessment of temperature issues, and/or the small size of the experimental groups considered in the "other potential threats to internal validity".

3.4.2.2. Fetal length. Eight studies were classified at "some concern" and 3 studies were classified at "high concern". The main reasons for "some concern" were limited confidence in the outcome assessment and lack of blinding during experiment performance. The main reasons for "high concern" were insufficient exposure characterization, and/or the small size of the experimental groups considered in the "other potential threats to internal validity".

3.4.2.3. Fetal malformations. Twelve studies were classified at "some concern", 7 studies were classified at "high concern" and 1 study was classified at "low concern". The main reasons for "some concern" were limited confidence in the outcome assessment and lack of blinding during experiment performance. The main reasons for "high concern" were insufficient exposure characterization, inadequate assessment of temperature issues, the small size of the experimental groups considered in the "other potential threats to internal validity" and/or inadequate comparators.

3.4.2.4. Litters with malformed fetuses. Twelve studies were classified at "some concern" and 2 studies were classified at "high concern". The main reasons for "some concern" were limited confidence in the outcome assessment and lack of blinding during experiment performance. The main reason for "high concern" was the small size of the experimental groups considered in the "other potential threats to internal validity".

3.4.2.5. Sex ratio. Eleven studies were classified at "some concern" and 1 study was classified at "low concern". The main reasons for concern were limited confidence in the outcome assessment and lack of blinding during experiment performance.

3.4.3. Delayed effects on the offspring health

3.4.3.1. Brain pathology. Nine studies were classified at "some concern" and 3 studies were classified at "high concern". The main reasons for "some concern" were limited confidence in the outcome assessment and lack of blinding during experiment performance. The reasons for "high concern" were inadequate comparators, insufficient exposure characterization and/or the small size of the experimental groups considered in the "other potential threats to internal validity".

3.4.3.2. Behavioural ontogeny. All 6 studies were classified at "high concern" because of lack of confidence in outcome assessment.

3.4.3.3. Learning and memory functions. Two studies were classified at "some concern" and 14 studies were classified at "high concern". The main reasons for "some concern" were lack of randomization or insufficient exposure characterization and inadequate assessment of temperature issues. The reasons for "high concern" were lack of confidence in outcome assessment, insufficient exposure characterization, inadequate comparators and/or the small size of the experimental groups considered in the "other potential threats to internal validity".

3.4.3.4. Motor activity functions. Three studies were classified at "some concern" and 11 studies were classified at "high concern". The main reasons for "some concern" were lack of blinding during experiment performance and/or lack of randomization. The reasons for "high concern" were lack of confidence in outcome assessment, insufficient exposure characterization and/or the small size of the experimental groups considered in the "other potential threats to internal validity".

3.4.3.5. Motor and sensory functions. One study was classified at "some concern" because of lack of blinding during the experiment performance and lack of randomization. One study was classified at "high concern" because of lack of confidence in outcome assessment.

3.4.3.6. Female infertility. Five studies were classified at "some concern" and 1 study was classified at "high concern". The main reasons for "some concern" were lack of blinding during experiment performance and limited confidence in the outcome assessment. The reason for "high concern" was the small size of the experimental groups considered in the "other potential threats to internal validity".

In general, the highest confidence in exposure characterization appeared to correlate with studies with the largest sample sizes: the 31 studies with "definitely low RoB" for exposure had a mean sample size of 64 (SD = 52), while the 11 studies with "definitely high RoB" for exposure had a mean sample size of 31 (SD = 23), which gives a mean difference of 33 (95% CI 1 to 65).

3.5. Effects of the exposure

3.5.1. Results of individual studies

Results of individual studies are reported in Supplementary File 7(a-e), along with experimental design and exposure conditions applied in each of them.

3.5.2. Results of the syntheses

3.5.2.1. Reduction of fecundity

3.5.2.1.1. Pre-implantation loss. A total of 5 studies reported data on pre-implantation loss. The only one rated at "low or some concern" RoB level yielded an SMD of −0.31 (95% CI −0.88 to 0.26) showing no impact of RF-EMF exposure on this endpoint. The pooled SMD of the other 4 studies rated at "high concern" RoB level was −0.36 (95% CI −0.73 to 0.00) (Fig. 3). Since only one study was rated at "some concern" RoB level neither subgroup nor dose-response analysis were performed.

3.5.2.1.2. Pre-implantation loss: Other studies not included in the meta-analysis. Three studies, evaluated at "some concern" RoB level, could not be included in the *meta*-analysis, because data on variation parameters were not provided (Lary et al., 1982, Lee et al., 2009) or because the effect was only expressed by pregnancy rate (Nawrot et al., 1985). Two of them showed no effect on pre-implantation loss after exposure levels of 2 or 4 W/kg for 90 min/day for 17 days without a substantial animal core temperature increase (Lee et al., 2009) or 12 W/kg for about 30 min for 1 day with about a 4 °C core temperature increase (Lary et al., 1982). The third study (Nawrot et al., 1985) showed a significant

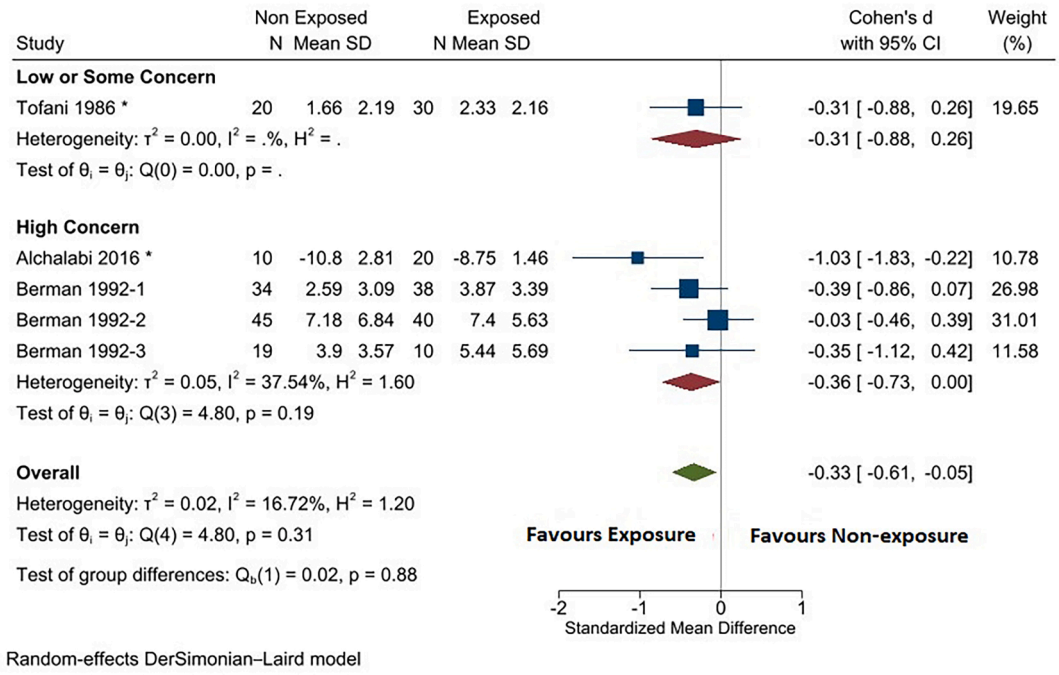

Fig. 3. Forest plot of studies on pre-implantation loss categorised as "low or some concern" or "high concern" for RoB. Tofani 1986 expressed the endpoint by the mean % losses per litter; Alchalabi 2016 expressed the endpoint by the mean number of implantation sites per litter; the 3 studies by Berman 1980 expressed the endpoint by the mean number of losses per litter. The bottom lines report the results and statistics of the *meta*-analysis for all included studies. Asterisks mark studies in which data from multiple exposure groups were combined to match a single comparator group. Progressive numbers after a reference indicate different studies reported in the same paper.

decrease in pregnancy rate after exposure to 40.2 W/kg for 480 min/day for 6 days and a 2.3 °C core temperature increase.

3.5.2.1.3. Litter size. A total of 36 studies reported data on litter size. The median litter size in the sham control groups across all 36 studies was 10.9 pups. In the 24 studies with "low or some concern" for overall study RoB, the mean difference between litter sizes in the sham-exposed and RF-EMF exposed groups was small and not statistically significant (MD 0.05 pup/litter, 95% CI –0.21 to 0.30). These studies were conducted at an average exposure level of 4.92 W/kg (SD 5.39, 0.02–22.2 min–max), with 42% testing a whole body average SAR equal to or higher than 5 W/kg. The pooled MD of studies at "high concern" RoB level was 0.77 (95% CI 0.15 to 1.39; 12 studies) (Fig. 4).

3.5.2.1.4. Litter size: Subgroup analysis. We could not use subgroup analyses by SAR and animal species for exploring causes of result heterogeneity because, in both cases, some of the subgroups included less than 3 studies. Subgrouping the studies by dam core temperature increase partly explained their heterogeneity. The mean difference of the 4 studies in which the dam core temperature increased over 1 °C (2 °C average increase; MD 0.99 pup/litter, 95% CI 0.36 to 1.62) was significantly different from the mean difference of the 9 studies in which this increase was not observed (MD –0.33 pup/litter, 95% CI –0.79 to 0.14) (Supplementary File 8).

3.5.2.1.5. Litter size: Dose-response analysis. The dose–response analysis, with a linear model fit, did not reveal a significant effect on litter size with each increase in exposure of 1 W/kg (MD 0.001 pup/litter, 95% CI –0.04 to 0.04; 24 studies, 31 observations). Based on the AIC there was no indication that a non-linear relation with cubic splines would better fit the data (Supplementary File 9).

3.5.2.1.6. Litter size: Other studies not included in the meta-analysis. Eight papers reporting results that could not be included in the *meta*-analysis were classified at "low concern" (1 paper), "some concern" (6 papers) or "high concern" (1 paper) for overall study RoB. In agreement with the result of the *meta*-analysis, studies from these papers did not show an effect of RF-EMF exposure on litter size at a range of exposure levels from 0.034 to 12 W/kg (Anderson et al., 2004, Bornhausen and Scheingraber 2000, Dasdag et al., 2000, Kaplan et al., 1982, Petitdant et al., 2018, Rifat et al., 2016, Shirai et al., 2017, Wyde et al., 2018). No effect on litter size was reported in the study by Cobb et al., 2000 where the effects of EMP were tested.

3.5.2.1.7. Resorbed or dead fetuses. A total of 27 studies reported data on the incidence of resorbed or dead fetuses. No study was rated at "low concern" RoB level. Studies classified at "some concern" for overall study RoB had an OR of 1.84 (95% CI 1.27 to 2.66; 21 studies), showing a statistically significant increase of resorbed or dead fetuses in the RF-EMF exposed animals. These studies were conducted at an average exposure level of 20.26 W/kg (SD 24.52, 0.0001–107.28 min–max), with 81% testing a whole body average SAR equal to or higher than 5 W/kg. Studies rated at "high concern" RoB level had an OR of 1.71 (95% CI 0.65 to 4.46; 6 studies) (Fig. 5).

3.5.2.1.8. Resorbed or dead fetuses: Subgroup analysis. Comparison of studies according to the animal species showed a borderline significant difference (p less than 0.07). We could not use the subgrouping by SAR for investigating causes of result heterogeneity because the group at the lowest SAR included only 1 study. Subgrouping studies by whether animal core temperature increased or not, did not show a significant difference between the 4 studies in which less than 1 °C increase was induced (OR 2.17, 95% CI 0.39 to 12.03) and the 14 studies in which the temperature of the exposed dams increased by over 1 °C (3 °C average increase; OR 1.59, 95% CI 1.15 to 2.19) (Supplementary File 8).

3.5.2.1.9. Resorbed or dead fetuses: Dose-response analysis. The linear fitting of the dose-response relationship showed a small but statistically significant increase of the OR of 0.03 per unit W/kg increase. The cubic spline model did not seem to fit the data better than the linear model (Supplementary File 9).

3.5.2.1.10. Resorbed or dead fetuses: Other studies not included in the meta-analysis. Two studies, rated at "some concern" for overall study RoB were not entered into the *meta*-analysis because data on variation parameters were not reported (Nelson et al., 2001), or data could not be extracted from the figures (Nelson et al., 1994). In agreement with the result of the *meta*-analysis, these studies showed an increased number of

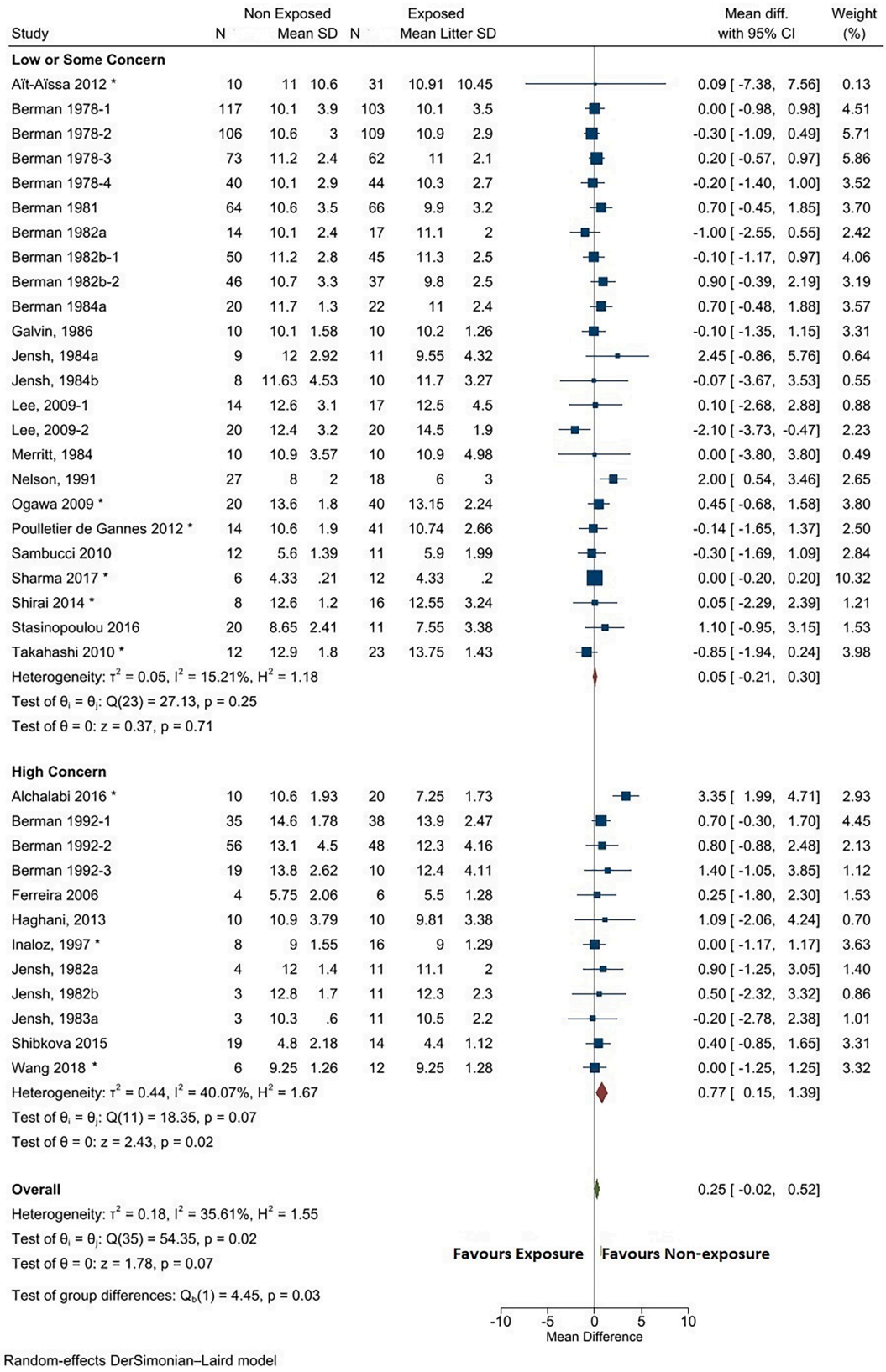

Fig. 4. Forest plot of studies on litter size categorised as "low or some concern" or "high concern" for RoB. The endpoint is expressed by the mean number of fetuses/pups per litter. The bottom lines report the results and statistics of the *meta*-analysis for all included studies. Asterisks mark studies in which data from multiple exposure groups were combined to match a single comparator group. Progressive numbers after a reference indicate different studies reported in the same paper.

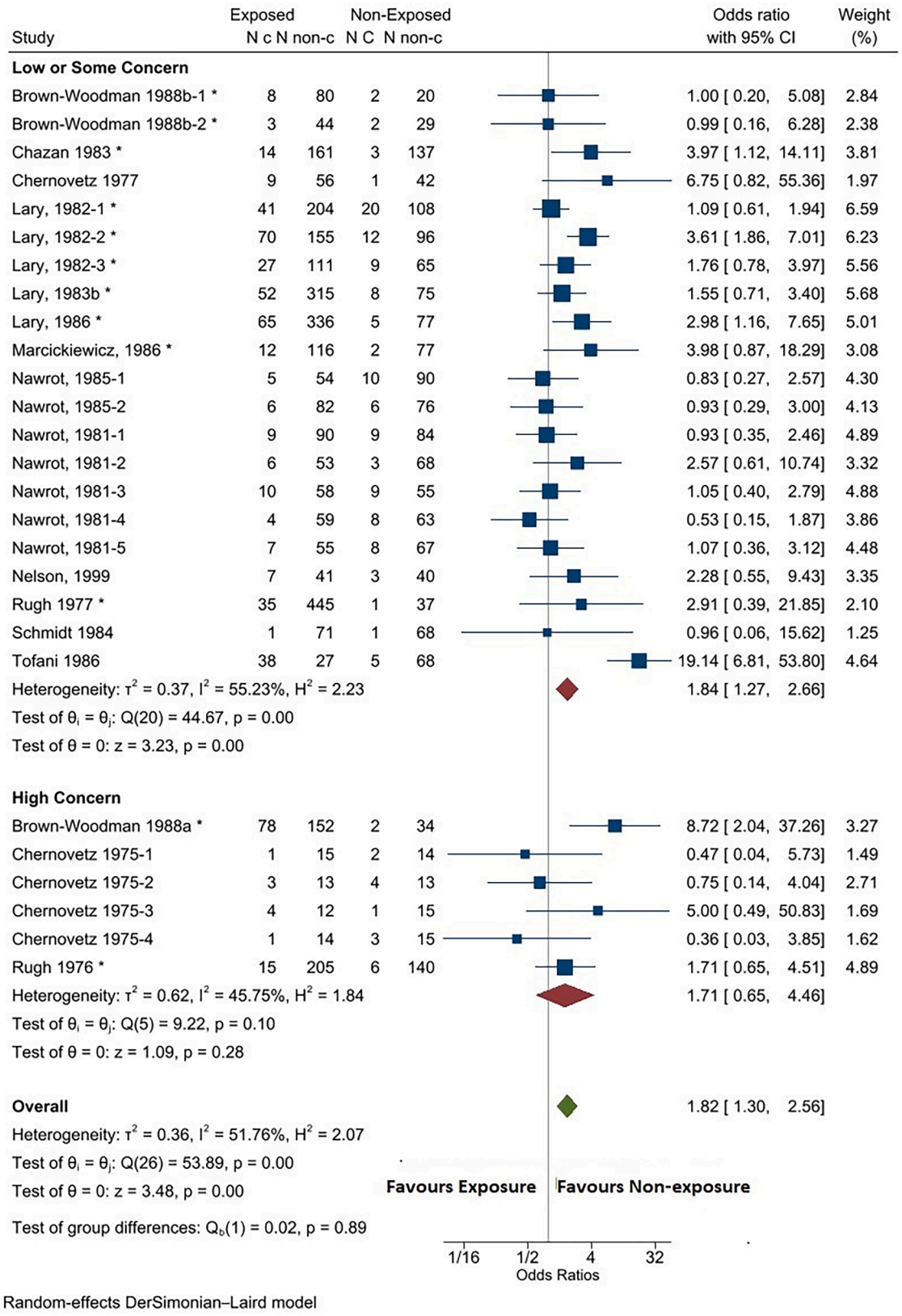

Fig. 5. Forest plot of studies on the incidence of resorbed or dead fetuses categorised as "low or some concern" or "high concern" for RoB. The endpoint is expressed by the number of resorbed or dead fetuses considered as "cases" (N c) vs the number of live fetuses (N non-c), after adjustment for intra-litter clustering. The bottom lines report the results and statistics of the *meta*-analysis for all included studies. Asterisks mark studies in which data from multiple exposure groups were combined to match a single comparator group. Progressive numbers after a reference indicate different studies reported in the same paper.

resorptions after high level RF-EMF exposures where core temperature increases of 3–4 °C were observed (Nelson et al., 1994, 2001).

3.5.2.2. *Adverse effects on the offspring health at birth*

3.5.2.2.1. *Fetal weight.*

Fig. 6 shows the forest plot of data on fetal weight categorized as "low or some concern" or "high concern" for RoB. Forty-eight studies rated as "low or some concern" had an SMD of 0.31 (95% CI 0.15 to 0.48) showing a small but statistically significant decrease of fetal weight in the RF-EMF exposed animals. These studies were conducted at an average exposure level of 9.83 W/kg (SD 11.85,

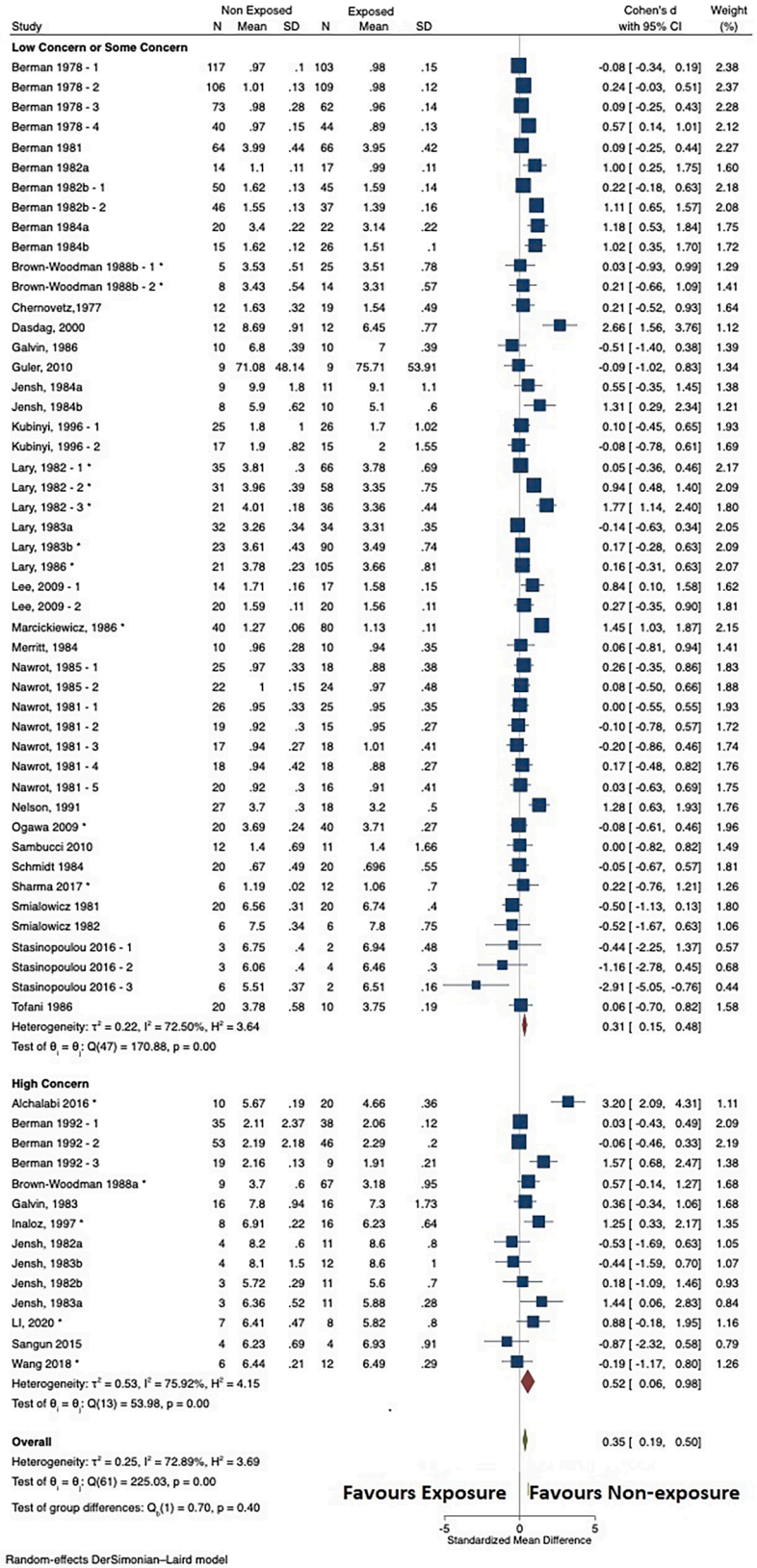

Fig. 6. Forest plot of studies on fetal weight categorised as "low or some concern" or "high concern" for RoB. The endpoint is expressed by the mean fetal weight in grams. SMD was used as the effect size measure because studies used various species that differed appreciably in weight. The bottom lines report the results and statistics of the *meta*-analysis for all included studies. Asterisks mark studies in which data from multiple exposure groups were combined to match a single comparator group. Progressive numbers after a reference indicate different studies reported in the same paper.

0.0001–40.2 min–max), with 54% testing a whole body average SAR equal to or higher than 5 W/kg. Fourteen studies rated at "high concern" RoB level had an SMD of 0.52 (95% CI 0.06 to 0.98).

3.5.2.2.2. Fetal weight: Subgroup analysis. Subgroup analysis according to animal species could not be used for investigating causes of result heterogeneity because of a paucity of studies for some species. After subgrouping by SAR, significantly different SMD values of −0.57 (95% CI −1.38 to 0.24, 5 studies), 0.07 (95% CI −0.15 to 0.29, 17 studies) and 0.51 (95% CI 0.31 to 0.72, 26 studies) were observed at whole body average SARs of less than 0.1, 0.1–5 and ≥ 5 W/kg, respectively. Subgrouping studies by whether animal core temperature increased or not, did not show significant differences between studies in which less than 1 °C increase was induced (SMD 0.36, 95% CI 0.11 to 0.60, 18 studies) and studies in which over 1 °C increase was induced (SMD 0.57, 95% CI 0.28 to 0.86, 16 studies) (Supplementary File 8).

3.5.2.2.3. Fetal weight: Dose response analysis. According to the dose response analysis with the linear model, there would be an increase in SMD of 0.015 with every unit increase of W/kg, which indicates a decrease of body weight in the RF-EMF exposed fetuses. According to the decrease of the AIC, the cubic spline analysis fitted the observed data better indicating a supralinear effect until about 7 W/kg (Supplementary File 9).

3.5.2.2.4. Fetal weight: Other studies not included in the meta-analysis. Seven papers reported on fetal weight assessment but did not provide data to be included in the *meta*-analysis or lacked essential information; six of them were classified at "some concern" and one at "high concern" for overall study RoB. Five of these papers did not observe a statistically significant fetal weight decrease in the RF-EMF exposed animals in spite of exposure at whole body average SAR as high as 7.9 W/kg and a dam core temperature increase as high as 3 °C (Chiang 1988, Nelson et al., 1999, 2001, Petitdant et al., 2018, Wyde et al., 2018). Two papers reported a significant decrease of fetal weight after a whole body average exposure to 6.6 W/kg RF-EMF that was associated with a dam core temperature increase of 4 °C (Nelson et al., 1994) or an exposure to 0.179 W/kg without a dam core temperature assessment (Rifat et al., 2016). In the paper by Cobb et al. (2000), effects of EMP were tested and no evidence of fetal weight decrease was reported.

3.5.2.2.5. Fetal length. A total of 15 studies reported data on fetal length. Thirteen studies rated as "some concern" for overall study RoB had a pooled SMD of 0.45 (95% CI 0.07 to 0.83), showing a small but statistically significant decrease of fetal length in RF-EMF exposed animals. These studies were conducted at an average exposure level of 4.55 W/kg (SD 5.35, 0.018–11.97 min–max), with 38% testing a whole body

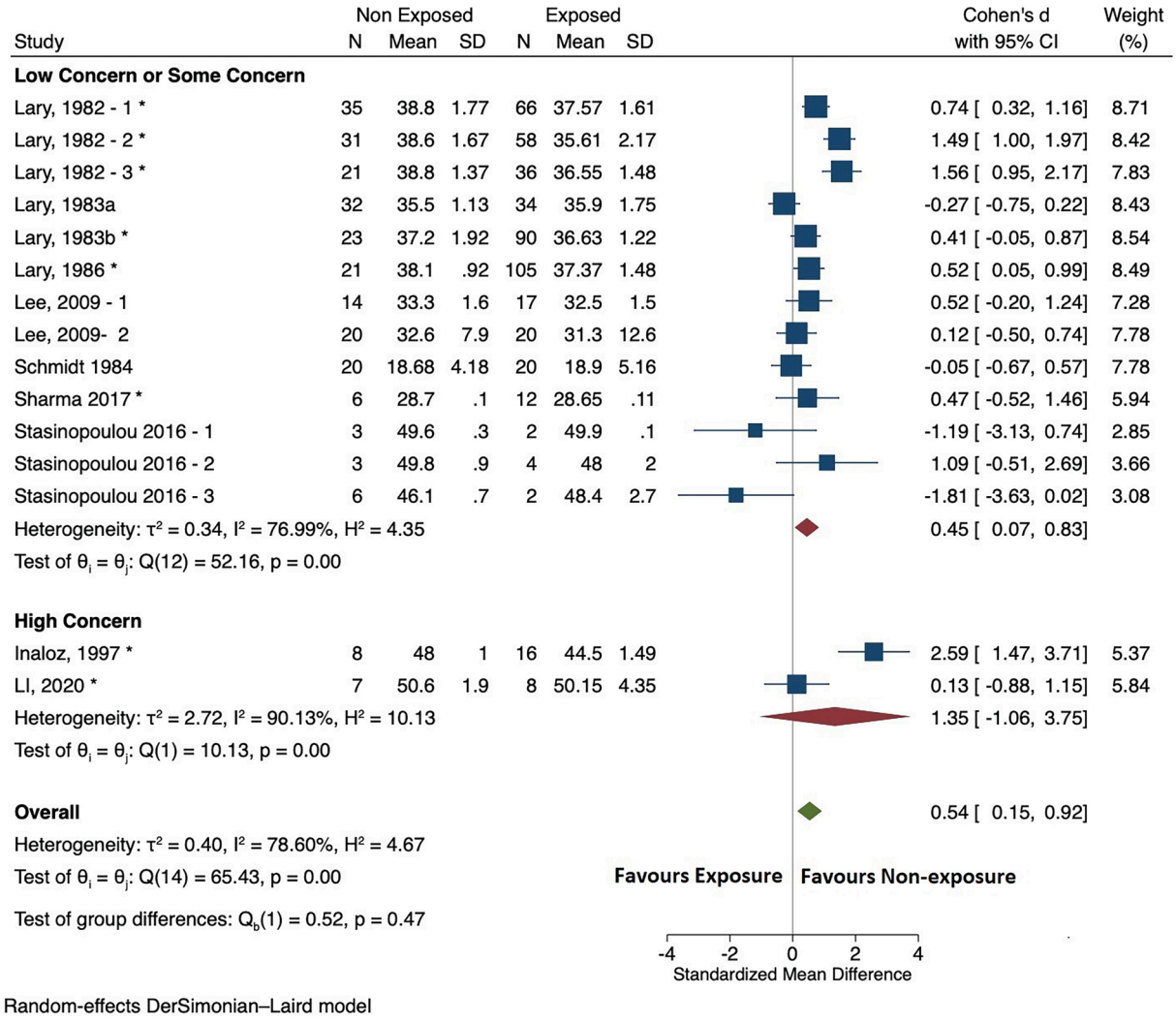

Fig. 7. Forest plot of studies on fetal length categorised as "low or some concern" or "high concern" for RoB. The endpoint is expressed by the mean fetal length in millimeters. SMD was used as the effect size measure because studies used various species that differed appreciably in length. The bottom lines report the results and statistics of the *meta*-analysis for all included studies. Asterisks mark studies in which data from multiple exposure groups were combined to match a single comparator group. Progressive numbers after a reference indicate different studies reported in the same paper.

average SAR equal to or higher than 5 W/kg. Studies rated at "high concern" RoB level had a pooled SMD of 1.35 (95% CI −1.06 to 3.75; 2 studies) (Fig. 7).

3.5.2.2.6. Fetal length: Subgroup analysis. Subgrouping by animal species did not show a significant difference between studies in mice and rats. Categorization of studies by exposure level showed a significant heterogeneity among groups. SMD values of −0.58 (95% CI −2.40 to 1.24, 3 studies), 0.05 (95% CI −0.24 to 0.34, 5 studies) and 0.92 (95% CI 0.47 to 1.37, 5 studies) were calculated for whole body average SARs below 0.1 W/kg, between 0.1 and 5 W/kg and above 5 W/kg subgroups, respectively. There was a significant difference between the pooled effect size of studies in which a temperature increase of less than 1 °C was reported (SMD 0.06, 95% CI −0.38 to 0.51, 3 studies) and the pooled effect size of studies in which a temperature increase equal to or higher

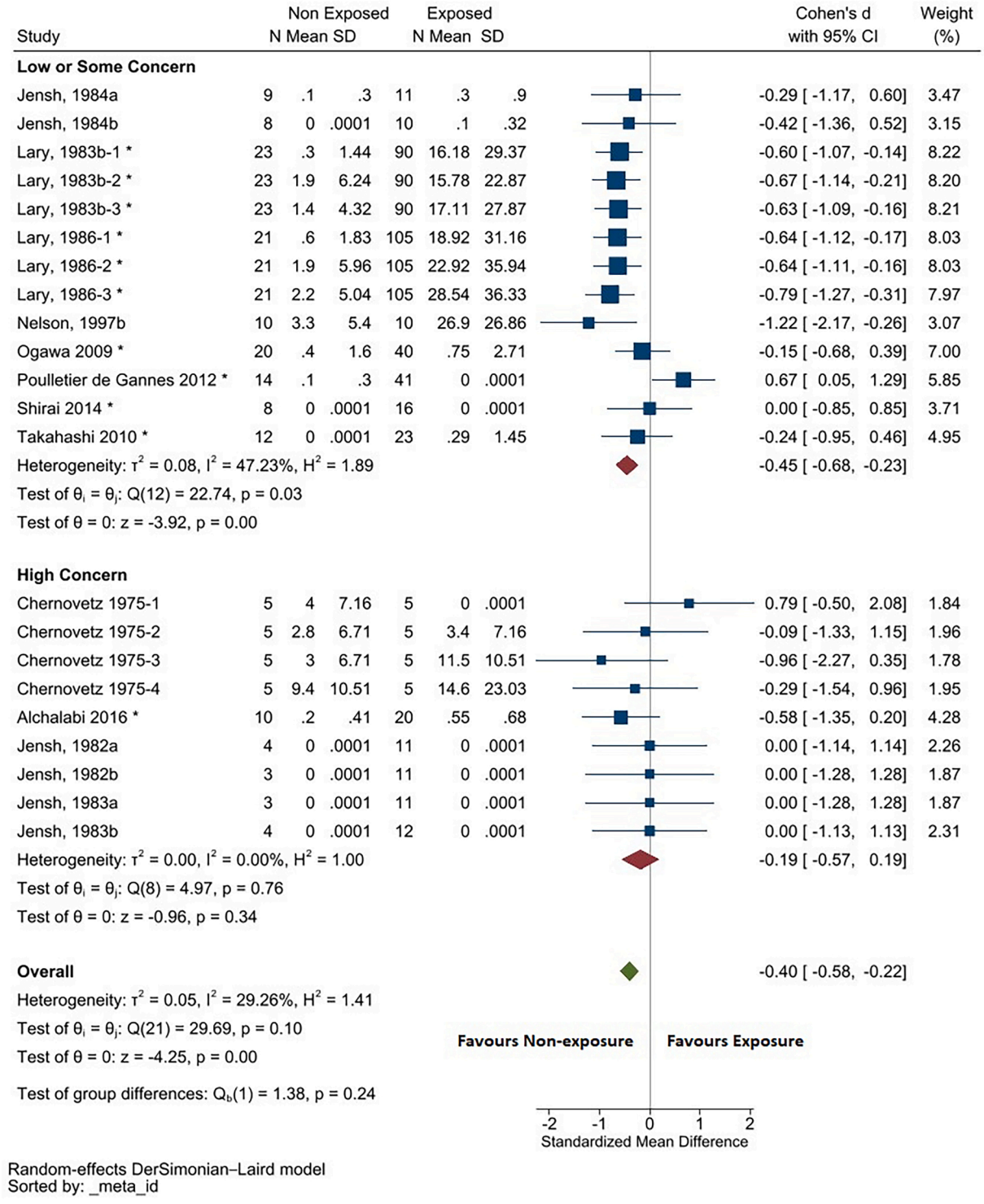

Fig. 8. Forest plot of studies on fetal malformations categorised as "low or some concern" or "high concern" for RoB. In some studies the endpoint is expressed by the mean number of malformations per litter, while in other studies the endpoint is expressed by the mean percentage of malformed fetuses per litter. The bottom lines report the results and statistics of the *meta*-analysis for all included studies. Asterisks mark studies in which data from multiple exposure groups were combined to match a single comparator group. Progressive numbers after a reference indicate different studies reported in the same paper.

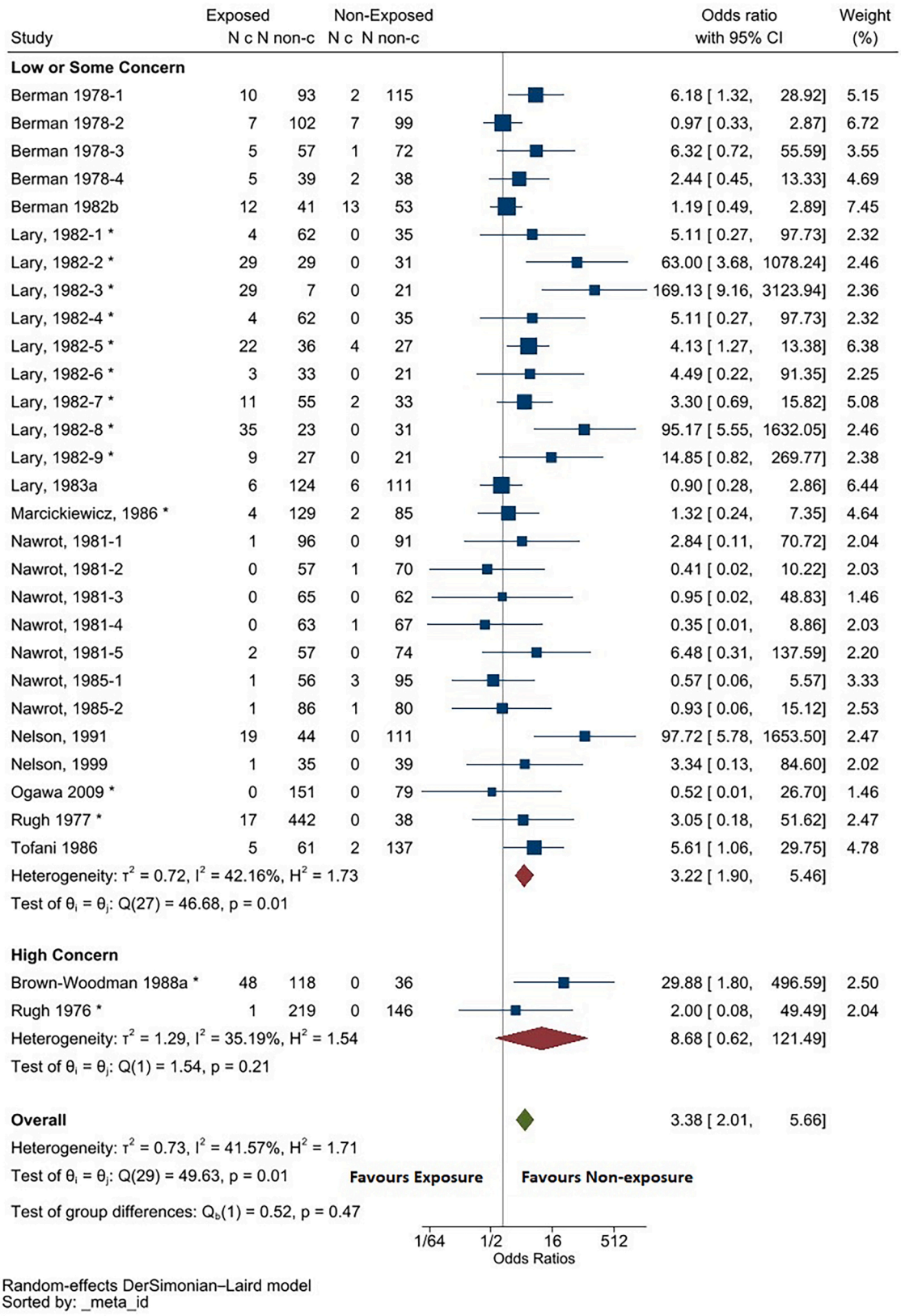

Fig. 9. Forest plot of studies on the incidence of litters with malformed fetuses categorised as "low or some concern" or "high concern" for RoB. In some studies the endpoint is expressed by the number of litters with at least one malformed fetus, as "cases" (N c) vs the number of litters without any malformed fetus (N non-c), while in other studies the endpoint is expressed by the number of malformed fetuses as "cases" (N c) vs the number of normal fetuses (N non-c). Numbers were adjusted for intra-litter clustering. The bottom lines report the results and statistics of the *meta*-analysis for all included studies. Asterisks mark studies in which data from multiple exposure groups were combined to match a single comparator group. Progressive numbers after a reference indicate different studies reported in the same paper.

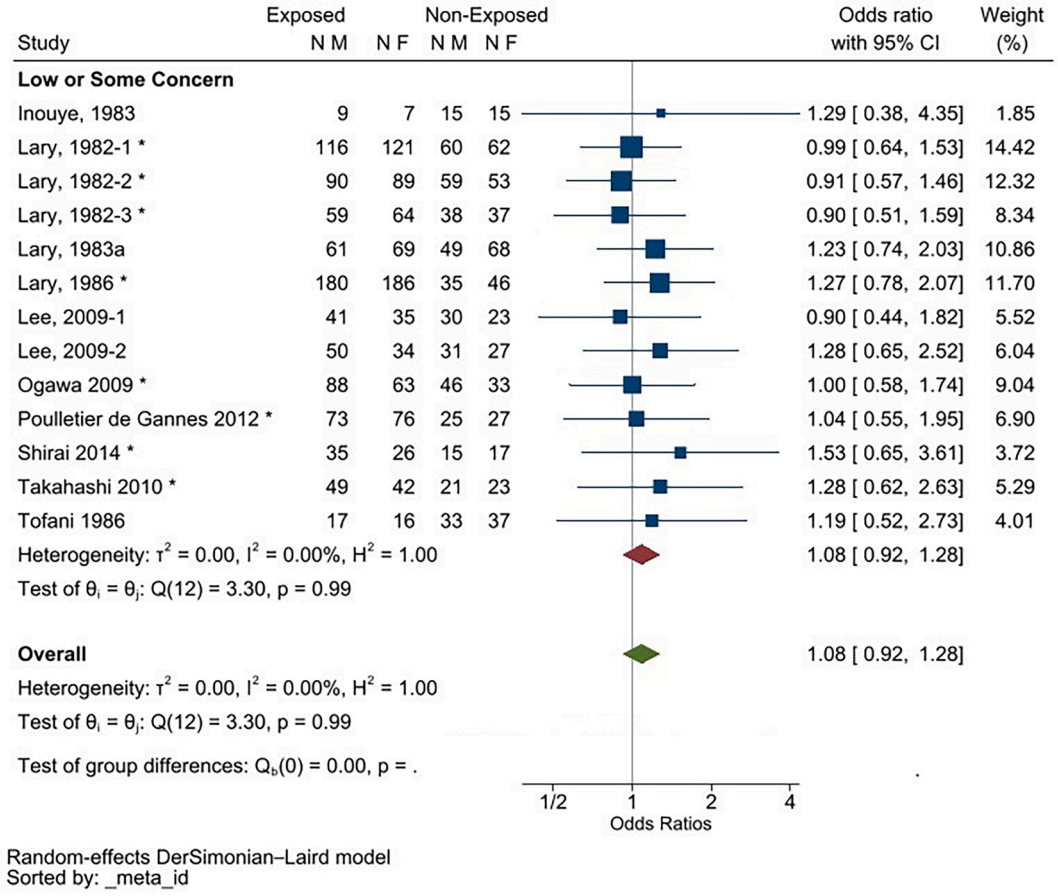

Fig. 10. Forest plot of studies on sex ratio categorised as "low or some concern" or "high concern" for RoB. The endpoint is expressed by the number of males as "cases" (N M) vs the number of females (N F), after adjustment for intra-litter clustering. The bottom lines report the results and statistics of the *meta*-analysis for all included studies. Asterisks mark studies in which data from multiple exposure groups were combined to match a single comparator group. Progressive numbers after a reference indicate different studies reported in the same paper.

than 1 °C was reported (SMD 0.92, 95% CI 0.47 to 1.37, 5 studies) (Supplementary File 8).

3.5.2.2.7. Fetal length: Dose-response analysis. The linear fitting showed a small but significant increase of the SMD of 0.03 for every W/kg unit increase, which indicates a decrease of body length in the RF-EMF exposed fetuses. The higher AIC value of the cubic spline fitting with respect to the linear one did not support a supralinear dose-response relationship (Supplementary File 9).

3.5.2.2.8. Fetal length: Other studies not included in the meta-analysis. The study described in Rifat et al. (2016), rated at "high concern" for overall study RoB, did not report evidence of an RF-EMF effect on fetal length after exposure to 0.179 W/kg.

3.5.2.2.9. Fetal malformations (continuous data). Twenty-two studies reported data on fetal malformations as a continuous variable. The *meta*-analysis of studies rated as "low or some concern" showed an SMD value of −0.45 (95% CI −0.68 to −0.23, 13 studies), showing a small but statistically significant increase of malformed fetuses in RF-EMF exposed animals. These studies were conducted at an average exposure level of 6.75 W/kg (SD 4.65, 0.048–10.8 min–max), with 69% testing a whole body average SAR equal to or higher than 5 W/kg. Studies rated at "high concern" RoB level had an SMD of −0.19 (95% CI −0.57 to 0.19; 9 studies) (Fig. 8).

3.5.2.2.10. Fetal malformations: Subgroup analysis. All studies classified as "low concern" or "some concern" RoB were performed in rats. The subgroup analyses by SAR and animal temperature increase could not be used to investigate causes of result heterogeneity because, in both cases, some of the subgroups included less than 3 studies (Supplementary File 8).

3.5.2.2.11. Fetal malformations: Dose-response analysis. The linear fitting showed a significant decrease of the SMD of −1.34 for every W/kg unit increase, which indicates an increase of malformations in the RF-EMF exposed fetuses. The similar AIC values between the linear and the cubic spline fittings do not suggest a supralinear model for the dose-response relationship (Supplementary File 9).

3.5.2.2.12. Litters with malformed fetuses (binary data). Some studies recorded malformation data as the number of litters with one or more malformed fetuses while other studies reported malformation data as the number of malformed fetuses over the total number of fetuses analysed. For a comprehensive *meta*-analysis, the latter set of data was adjusted for litter clustering and assimilated into the first set of data. Fig. 9 shows the forest plot of data categorized as "low or some concern" or "high concern" for RoB. The 28 studies rated at "some concern" for overall study RoB yielded an OR of 3.22 (95% CI 1.90 to 5.46), showing a significant increase of litters with malformed fetuses in the RF-EMF exposed animals, in spite of a large variability of individual study ORs from 0.35 to 169.13. These studies were conducted at an average exposure level of 16.63 W/kg (SD 20.96, 0.0001–107.28 min–max), with 86% testing a whole body average SAR equal to or higher than 5 W/kg. Studies rated at "high concern" RoB level had an OR of 8.68 (95% CI 0.62 to 121.49; 2 studies).

3.5.2.2.13. Litters with malformed fetuses: Subgroup analysis. Based on the number of studies in each subgroup, subgrouping by dam core temperature increase was the only analysis that could be conducted. OR values in studies where a temperature increase below or above 1 °C was detected were 2.01 (95% CI 0.96 to 4.21; 6 studies) and 4.53 (95% CI 2.08 to 9.85; 19 studies), respectively, but were not significantly

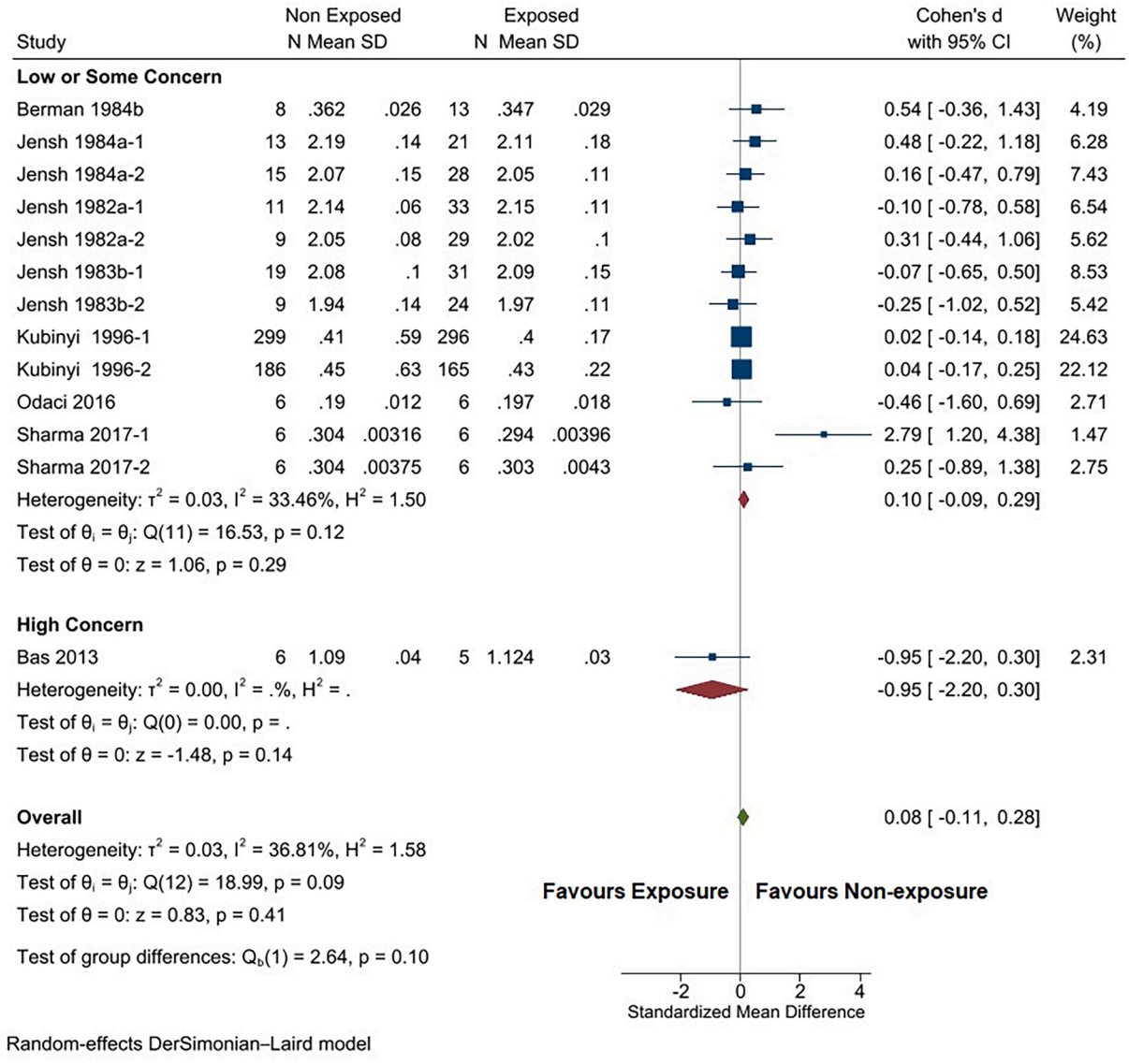

Fig. 11. Forest plot of studies on offspring brain weight categorised as "low or some concern" or "high concern" for RoB. The endpoint is expressed by brain weight in grams. SMD was used as the effect size measure because studies used various species that differed appreciably in brain weight. The bottom lines report the results and statistics of the *meta*-analysis for all included studies. Progressive numbers after a reference indicate different studies reported in the same paper.

different (Supplementary File 8).

3.5.2.2.14. Litters with malformed fetuses: Dose-response analysis. The linear and the cubic spline dose–response curves had similar AIC values (250.66 and 240.5, respectively). The linear fitting showed a small but significant increase of the OR for every W/kg unit increase indicative of increase of the incidence of litters with malformed fetuses in the RF-EMF exposed dams (Supplementary File 9).

3.5.2.2.15. Fetal malformations: Other studies not included in the meta-analyses. Eight studies described results that could not be included in the *meta*-analyses. The studies by Shirai et al. (2017), Berman et al. (1981, 1984a) and Lee et al. (2009) (rated at "some concern" for overall study RoB) reported no RF-EMF effect on fetal malformations up to exposure levels inducing a 2 °C animal core temperature increase. Nelson et al. (1994, 1997a, 2001) (rated at "some concern" for overall study RoB) reported an effect on fetal malformations at an RF-EMF exposure level that induced a dam core temperature increase above 1.5 °C. Alchalabi et al. (2017) (rated at "high concern" for overall study RoB) reported variable effects on skeletal development and malformations at a whole body average SAR exposure level of 0.974 W/kg.

3.5.2.2.16. Sex ratio. The *meta*-analysis of 13 studies, all rated as "low or some concern" yielded an OR of 1.08 (95% CI 0.92 to 1.28), showing no effect of RF-EMF on this endpoint (Fig. 10). These studies were conducted at an average exposure level of 4.29 W/kg (SD 5.093, 0.0001–10.8 min–max), with 31% testing a whole body average SAR equal to or higher than 5 W/kg.

3.5.2.2.17. Sex ratio: Subgroup analysis. There was no difference among subgroups either by SAR level or animal temperature increase. Subgrouping by animal species could not be used because there were too few studies in mice (Supplementary File 8).

3.5.2.2.18. Sex ratio: Dose-response analysis. No dose–response relationship analysis was conducted because no individual study tested more than one exposure level.

3.5.2.2.19. Sex ratio: Other studies not included in the meta-analysis. The study by Shirai et al. (2017) ("some concern" for overall study RoB) did not report data suitable for inclusion in the *meta*-analysis. Similar to the *meta*-analysis result, no effects were observed in this study after SAR exposure levels of 0.087 and 0.433 W/kg for 20 h/day for 16 days. The study by Cobb et al. (2000) that tested EMP did not detect an effect of exposure.

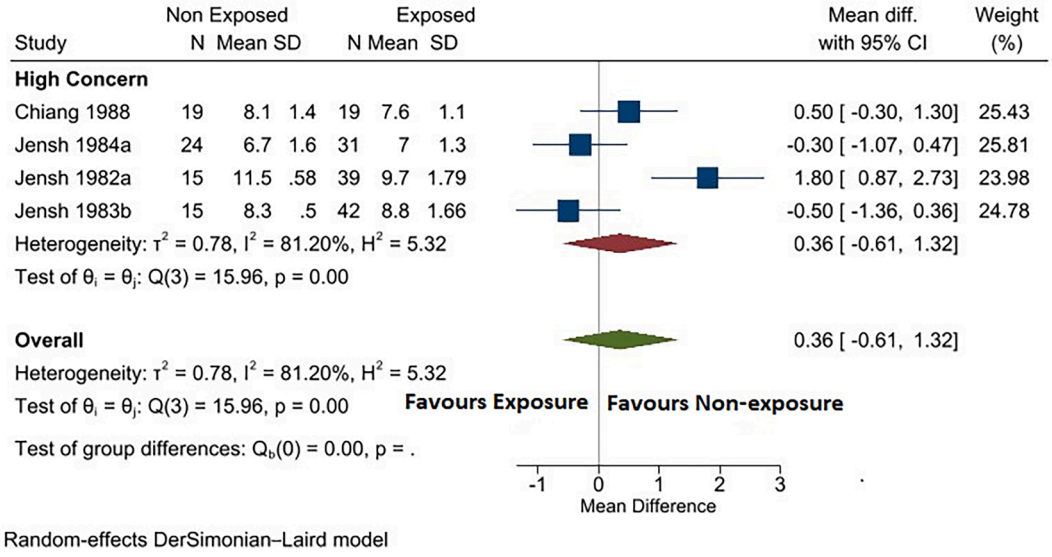

Fig. 12. Forest plot of studies on age of first righting, as behavioural ontogeny biomarker, all at "high concern" for RoB. The endpoint is expressed by the post-natal day of first righting. The bottom lines report the results and statistics of the *meta*-analysis for all included studies.

3.5.2.3. Delayed effects on the offspring health

3.5.2.3.1. Brain pathology. The weight of the brain or the cerebellum was considered the most representative marker of a possible RF-EMF impact on the central nervous system. These data were reported in a total of 13 studies. Twelve studies rated at "some concern" for RoB yielded an SMD of 0.10 (95% CI −0.09 to 0.29), showing no effects of

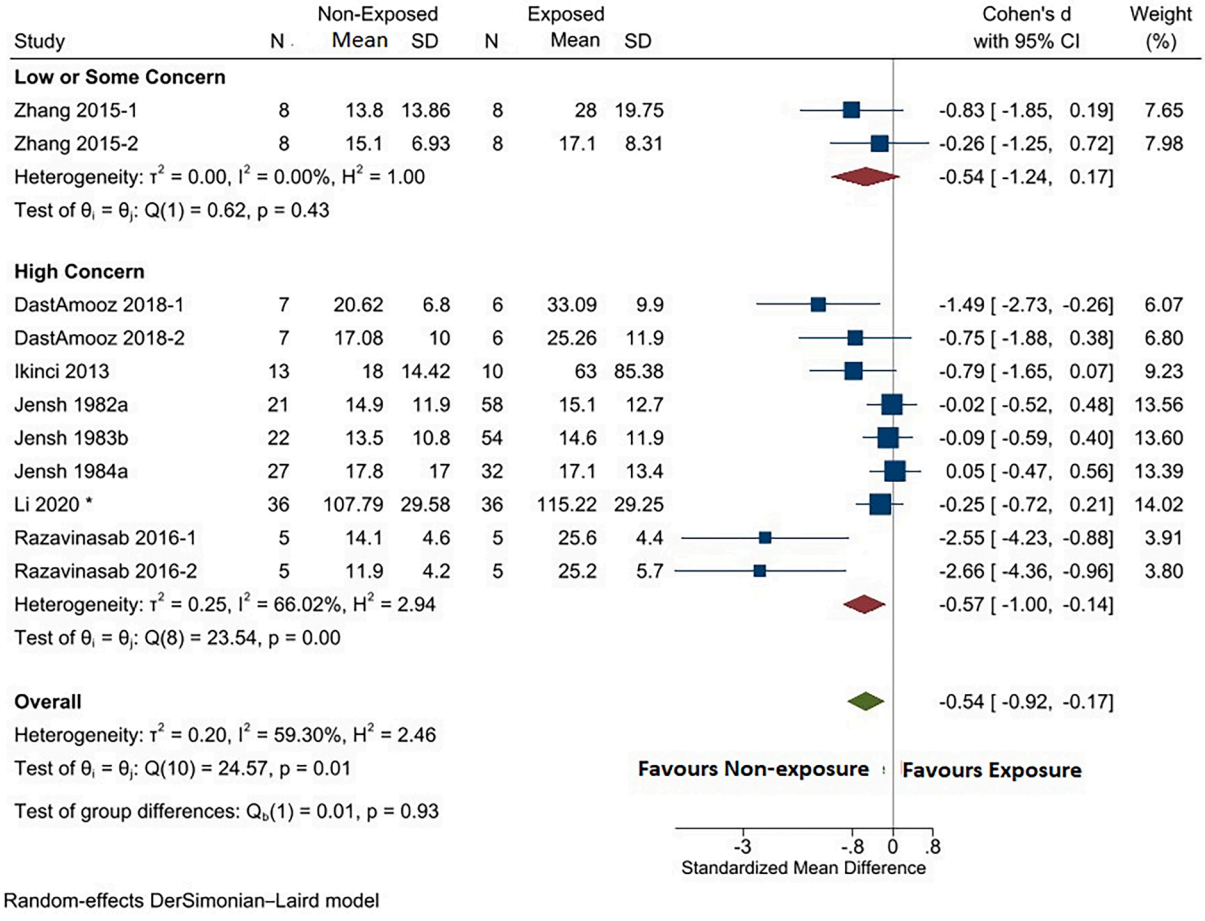

Fig. 13. Forest plot of studies on maze escape latency time, as learning and memory function biomarker, categorised as "low or some concern" or "high concern" for RoB. The endpoint is expressed by seconds needed to find an escape in a maze test. Since different types of mazes were tested, SMD was used as the effect size measure. The bottom lines report the results and statistics of the *meta*-analysis for all included studies. Asterisks mark studies in which data from multiple exposure groups were combined to match a single comparator group. Progressive numbers after a reference indicate different studies reported in the same paper.

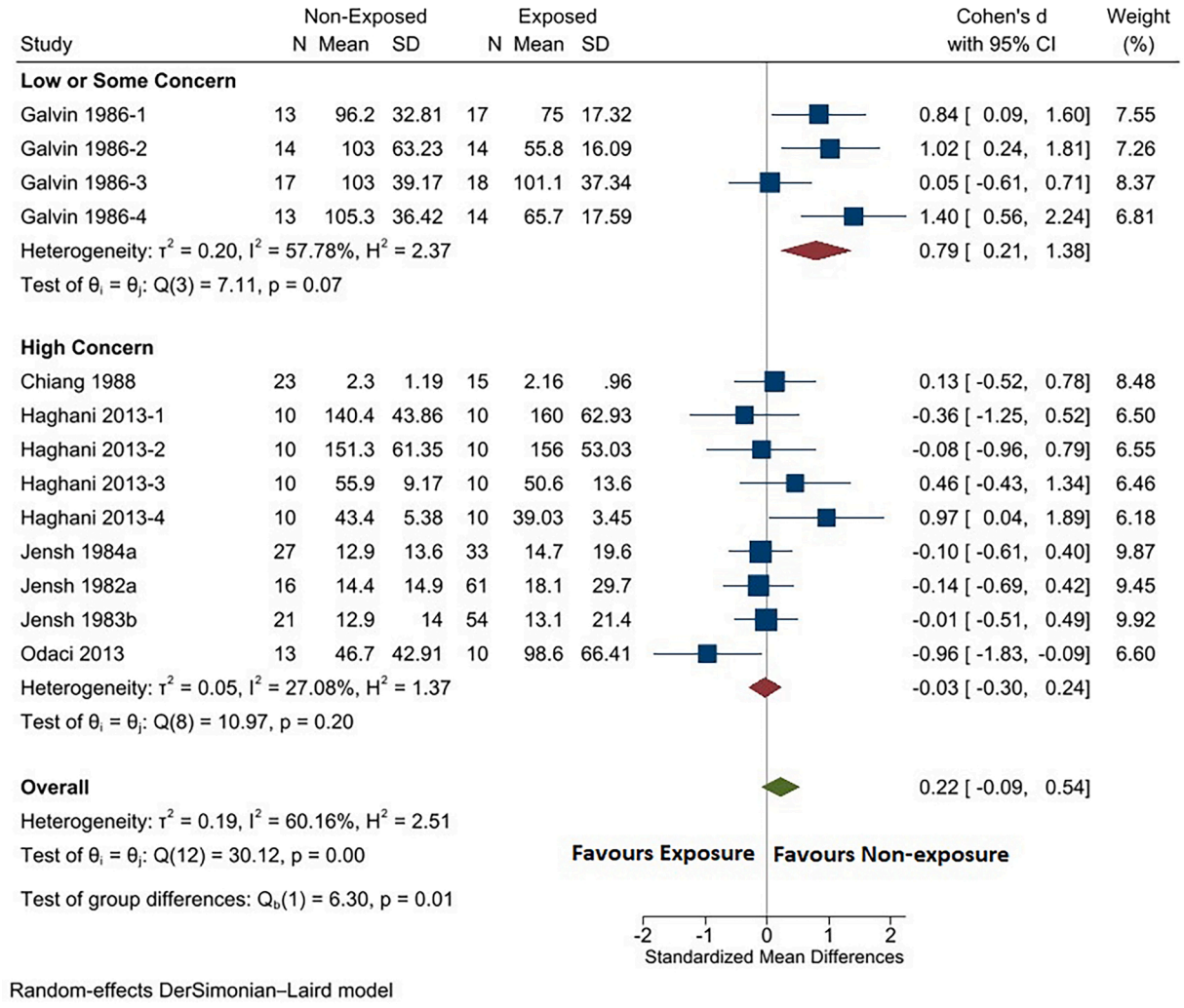

Fig. 14. Forest plot of studies on endurance time, as motor activity function biomarker, categorised as "low or some concern" or "high concern" for RoB. The endpoint is expressed by the time of endurance in seconds in motor activity tests. Since different types of tests were applied, SMD was used as the effect size measure. The bottom lines report the results and statistics of the *meta*-analysis for all included studies. Progressive numbers after a reference indicate different studies reported in the same paper.

RF-EMF exposure on this endpoint. The study rated at "high concern" RoB level had an SMD of −0.95 (95% CI −2.20 to 0.3) (Fig. 11).

No dose–response relationship analysis was conducted because no individual study tested more than one exposure level.

3.5.2.3.2. Brain pathology: Subgroup analysis. There was no statistically significant difference between the studies carried out in rats and mice. No subgroup analysis was conducted for SAR because some groups included less than 3 studies. All studies in which animal core temperature was measured reported an increase of less than 1 °C (Supplementary File 8).

3.5.2.3.3. Brain pathology: Other studies not included in the meta-analysis. In addition to brain or cerebellum weight, 7 papers reported data on the number or density of neural cells in specific portions of the central nervous system (Albert et al., 1981, Bas et al., 2013, Erdem Koç et al., 2016, Keles and Sut, 2021, Odaci et al., 2016, Sharma et al., 2017, Stasinopoulou et al., 2016). Four studies were classified at "some concern" and 3 at "high concern" for overall study RoB. All these papers reported a significant decrease in cell number or cell density at whole body average SARs ranging between 0.01 and 2 W/kg. When organ weight was also measured, two papers reported no impact on brain (Bas et al., 2013) or on cerebellum (Odaci et al., 2016), while 1 paper (Sharma et al., 2017) reported a decrease of brain weight. Considering the difference in the methods used to assess brain pathology, the difference between the results of the *meta*-analysis on brain weight and the results of the studies measuring neuronal cell numbers is not a surprise.

3.5.2.3.4. Behavioural ontogeny. All the 4 studies on the age of first righting included in the *meta*-analysis were rated at "high concern" RoB level and yielded an MD of 0.36 (95% CI −0.61 to 1.32) (Fig. 12). For this reason, no further subgrouping was conducted, and this set of data was not included in the assessment of the body of evidence.

3.5.2.3.5. Behavioural ontogeny: Other studies not included in the meta-analysis. The study by O'Connor (1988), in which animals were exposed for 6 h per day during the whole gestation to a power density of about 300 W/m^2, did not observe an RF-EMF impact on this marker. Similarly, no effect on air righting was reported by Cobb et al. (2000) after testing EMP.

3.5.2.3.6. Learning and memory functions. The escape latency time from a maze was considered the most representative biomarker of learning and memory functions. Fig. 13 shows the forest plot of data categorized as "low or some concern" or "high concern" for RoB. The 2 studies classified at "some concern" for RoB did not show a significant impact of RF-EMF exposure (SMD −0.54, 95% CI −1.24 to 0.17). No subgrouping analysis was applied to these studies, and no dose–response fitting was explored since the 2 studies tested the same exposure level. Studies rated at "high concern" RoB level had an SMD of −0.57 (95% CI

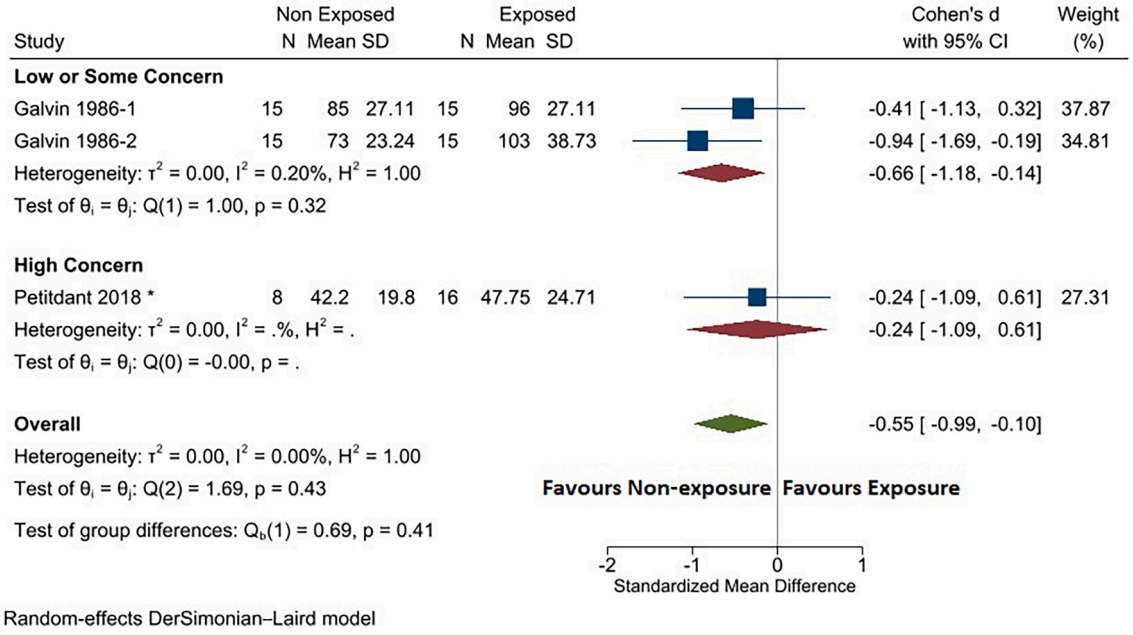

Fig. 15. Forest plot of studies on magnitude of startle to stimulus, a motor and sensory function biomarker, categorised as "low or some concern" or "high concern" for RoB. The endpoint is expressed in arbitrary units. The bottom lines report the results and statistics of the *meta*-analysis for all included studies. Asterisks mark studies in which data from multiple exposure groups were combined to match a single comparator group. Progressive numbers after a reference indicate different studies reported in the same paper.

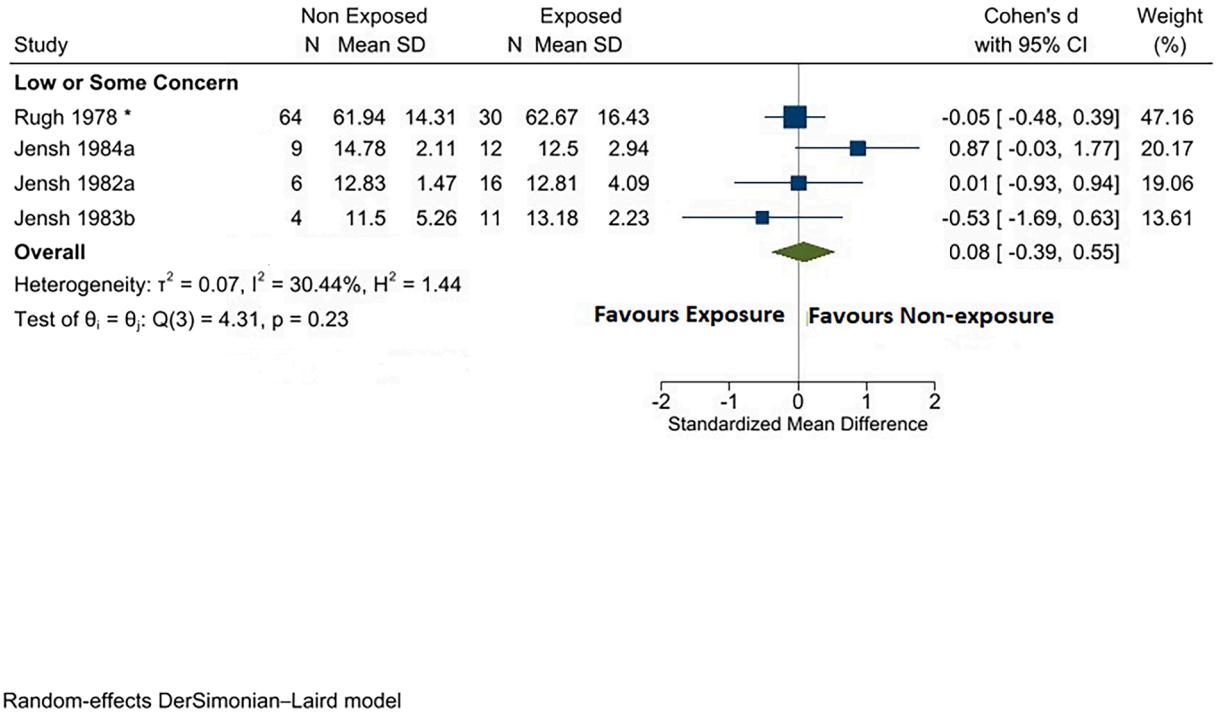

Fig. 16. Forest plot of studies on female offspring fertility, categorised as "low or some concern" for RoB. Rugh 1978 expressed the endpoint by the mean total number of pups in a series of subsequent litters, while, in the other studies, the endpoint was expressed by the mean number of pups per litter. The bottom lines report the results and statistics of the *meta*-analysis for all included studies. Asterisks mark studies in which data from multiple exposure groups were combined to match a single comparator group.

−1.00 to −0.14; 9 studies).

3.5.2.3.7. Learning and memory functions: Other studies not included in the meta-analysis. In addition to the studies that assessed the escape latency time, other studies applied a variety of behavioural markers reflecting learning and memory functions, with inconsistent results. The performance in a standard object recognition memory test was applied by Aldad et al. (2012) ("high concern" for RoB) who detected an impairment of the memory function after 24 h per day exposure to 1.6 W/kg. The passive avoidance learning and memory test was applied by Azimzadeh and Jelodar (2020), Ikinci et al. (2013) and Razavinasab et al. (2016) ("high concern" for RoB in all 3 papers). The 3 papers reported an RF-EMF effect on this behavioural marker after exposure to

Table 4
Summary of pooled effect sizes of "low or some" and "high" concern RoB studies for each specific endpoint.

Endpoint (effect size measure)	Pooled effect sizes of "low or some concern" RoB studies (95% CI) (N° studies)	Pooled effect sizes of "high concern" RoB studies (95% CI) (N° studies)
Reduction of fecundity		
Pre-implantation loss (SMD)	1 study only	−0.36 (−0.73 to 0) (4)
Litter size (MD)	0.05 (−0.21 to 0.3) (24)	0.77 (0.15 to 1.39) (12)
N° resorbed or dead fetuses (OR)	1.84 (1.27 to 2.66) (21)	1.71 (0.65 to 4.46) (6)
Adverse effects on the offspring health at birth		
Fetal weight (SMD)	0.31 (0.15 to 0.48) (48)	0.52 (0.06 to 0.98) (14)
Fetal length (SMD)	0.45 (0.07 to 0.83) (13)	1.35 (−1.06 to 3.75) (2)
Fraction of malformed fetuses (SMD)	−0.45 (−0.68 to −0.23) (13)	−0.19 (−0.57 to 0.19) (9)
N° litters with malformed fetuses (OR)	3.22 (1.9 to 5.46) (28)	8.68 (0.62 to 121.49) (2)
Sex ratio (OR)	1.08 (0.92 to 1.28) (13)	No study
Delayed effects on the offspring health		
Brain pathology (SMD)	0.10 (−0.09 to 0.29) (12)	1 study only
Behavioural ontogeny (MD)	No study	0.36 (−0.61 to 1.32) (4)
Learning and memory functions (SMD)	−0.54 (−1.24 to 0.17) (2)	−0.57 (−1.0 to −0.14) (9)
Motor activity (SMD)	0.79 (0.21 to 1.38) (4)	−0.03 (−0.3 to 0.24) (9)
Motor and sensory functions (SMD)	−0.66 (−1.18 to −0.14) (2)	1 study only
Female offspring fertility (SMD)	0.08 (−0.39 to 0.55) (4)	No study

whole body average SARs of 0.035, 0.4 and 0.6 W/kg, respectively. Jensh et al. (1983b) ("high concern" for RoB) did not detect an effect from exposure to 4.4 W/kg for 8 h per day for 14 days when assessed by shuttle box conditioned avoidance response test. Bornhausen and Scheingraber (2000) ("some concern" for RoB) applied a food reinforcement learning test in the offspring of dams exposed to 0.046 W/kg, 24 h per day, during the whole gestation and reported no evidence of cognitive impairment. No effect on learning and memory by freeze post-conditioning test was reported by Petitdant et al. (2018) ("high concern" for RoB) who exposed dams to 0.7 or 2.6 W/kg, 45 min per day, during the whole gestation. Finally, 3 papers reported data on the learning and memory function of the offspring by their performance in a maze test measured by parameters different from the escape latency time. Chernovetz et al. (1975) ("high concern" for RoB) did not detect RF-EMF effects after a short-duration (10 min) exposure on one day at 38 W/kg. Chiang (1988) ("high concern" for RoB) did not detect effects after whole gestation exposure to 3.25 W/kg for 5 h/day. Zhao et al. (2005) ("high concern" for RoB) tested 4 different frequencies, ranging between 37200 and 60000 MHz, at 4 different exposure levels up to 3.8 W/kg for 2 h per day for 10 days with variable effects increasing with the exposure level. Similar to the studies included in the *meta*-analysis, little confidence can be given to data from these studies because all but one of them were of "high concern" for RoB and due to the inconsistency of results across studies. Cobb et al. (2000) ("high concern" for RoB) did not observe any effect of EMP exposure on the offspring latency time in a water maze test.

3.5.2.3.8. Motor activity functions. The endurance time in any type of motor activity test was considered the most representative biomarker of motor activity functions. We included 13 studies in the *meta*-analysis. Fig. 14 shows the forest plot of data categorized as "low or some concern" or "high concern" for RoB. The 4 studies classified at "some concern" for RoB yielded an SMD of 0.79 (95% CI 0.21 to 1.38), showing a decrease of the endurance capacity in the progeny of the RF-EMF exposed dams. No subgrouping analysis and no dose–response fitting were applied to the 4 studies rated at "some concern" RoB level, since all of them came from the same paper and tested the same exposure level in the same experimental species. The studies rated at "high concern" RoB level had an SMD of −0.03 (95% CI −0.3 to 0.24; 9 studies).

3.5.2.3.9. Motor activity functions: Other studies not included in the meta-analysis. Eleven papers, 1 at "some concern" and 10 at "high concern" for RoB, reported data using the open field test. Ten did not observe an effect of RF-EMF exposure at levels as high as 7.28 W/kg (DastAmooz et al., 2018, Haghani et al., 2013, Jensh et al., 1982a, 1983b, Jensh 1984a, Kaplan et al., 1982, Li et al., 2020, O'Connor 1988, Odaci et al., 2013, Zhang et al., 2015). One of these papers used squirrel monkeys as the experimental animals (Kaplan et al., 1982) and also reported no effect on this endpoint. Only one paper, which tested 0.7 or 2.6 W/kg exposure for 45 min per day during the whole gestation, reported an RF-EMF effect at the highest exposure level, but only when the offspring were adolescents. The effect was no longer observed when the animals reached adult age (Petitdant et al., 2018). Since most of these studies were of "high concern" for RoB, these studies are not very informative. No effect of EMP exposure on the offspring open field performance was detected by Cobb et al. (2000).

3.5.2.3.10. Motor and sensory functions. Only 2 papers, including a total of 3 studies, analysed motor and sensory functions by magnitude of the startle response to stimuli. Fig. 15 shows the forest plot of data categorized as "low or some concern" or "high concern" for RoB. The pooled effect size of the 2 studies rated at "low or some concern" for RoB of −0.66 (95% CI −1.18 to −0.14) indicated a moderate increase of the magnitude of the startle response in the RF-EMF exposed animals. The study rated at "high concern" RoB level had an SMD of −0.24 (95% CI −0.99 to −0.10).

3.5.2.3.11. Female infertility. Fig. 16 shows the forest plot of data on F2 litter size. All studies were rated at "some concern" for RoB and did not show an effect of RF-EMF on this parameter (SMD 0.08, 95% CI −0.39 to 0.55; 4 studies).

3.5.2.3.12. Female infertility: Other studies not included in the meta-analysis. In addition to the studies measuring F2 litter size, 2 papers reported data about the follicle numbers in the offspring (Calis et al., 2019, Turedi et al., 2016). One study was rated at "some concern" and the other at "high concern" for RoB. Both studies reported significant decreases of follicle numbers in animals exposed *in utero* to a whole body average SAR of 0.23 and 0.01 W/kg, respectively.

3.5.3. Summary of results

Table 4 summarizes the pooled effect sizes measured for each endpoint. The values relative to the group of "low or some concern" studies and the group of "high concern" studies are separately reported. For some of the most relevant endpoints, like litter size and incidence of dead fetuses, it is shown that the pooled effect size of studies at "high concern" for RoB is much greater or much more uncertain than the value for studies at "low or some concern" for RoB. In other cases, this is not so evident, but this may be due to the small number of studies. These observations support the assumption that studies at high risk of bias report biased and exaggerated results or are more variable, which reduces the robustness of the pooled effect size. These considerations support our decision of using only the set of studies at "low or some concern" for RoB

Table 5
GRADE Evidence Profile.

Certainty assessment								Summary of findings				Certainty	Importance**
								No of participants		Effect			
No of studies	Design	RoB	Inconsistency	Indirectness	Imprecision	Publication bias *	Consistency across species	Exposure	Comparator	Relative(95% CI)	Absolute(95% CI)		
Reduction of fecundity													
Pre-implantation loss: No *meta*-analysis was done because the database included only one study that was not at "high concern" for RoB													6
Litter size[a] (an MD positive value indicates a detrimental RF-EMF effect)													
24	CES***	-1	0	0	0	0	+1	994	908		MD 0.05 pups(–0.21 to 0.3)	High	8
Resorbed or dead fetuses[b]													
21	CES***	-1	-1	0	0	0	0	3042	1569	OR 1.84(1.27 to 2.66)		Low	8
Adverse effects on the offspring health at birth													
Fetal weight[c] (an SMD positive value indicates a detrimental RF-EMF effect)													
48	CES***	-1	-1	0	0	0	+1	1477	1197	SMD 0.31(0.15 to 0.48)		Moderate	7
Fetal length[d] (an SMD positive value indicates a detrimental RF-EMF effect)													
13	CES***	-1	-1	0	0	0	0	466	235	SMD 0.45(0.07 to 0.83)		Low	7
Fetal malformations[e] (an SMD negative value indicates a detrimental RF-EMF effect)													
13	CES***	-1	0	0	0	-1	0	712	213	SMD -0.45(-0.68 to -0.23)		Low	8
Litters with malformed fetuses[f]													
28	CES***	-1	-1	-1	0	-1	0	2374	1818	OR 3.22(1.90 to 5.46)		Very low	8
Sex ratio[g]													
13	CES***	-1	0	-1	0	0	0	1719	948	OR 1.08(0.92 to 1.28)		Low	2
Delayed effects on the offspring health													
Brain pathology[h] (an SMD positive value indicates a detrimental RF-EMF effect)													
12	CES***	-1	0	0	0	0	0	658	587	SMD 0.10(-0.09 to 0.29)		Moderate	6
Behavioural ontogeny: No *meta*-analysis was done because the database included only studies at "high concern" for RoB													6
Learning and memory functions[i] (an SMD negative value indicates a detrimental RF-EMF effect)													
2	CES***	-1	0	0	-2	-1	0	16	16	SMD -0.54(-1.24 to 0.17)		Very low	7
Motor activity functions[j] (an SMD positive value indicates a detrimental RF-EMF effect)													
4	CES***	-1	-1	0	-1	0	0	63	57	SMD 0.79(0.21 to 1.38)		Very low	7
Motor and sensory functions[k] (an SMD negative value indicates a detrimental RF-EMF effect)													
2	CES***	-1	0	0	-2	0	0	30	30	SMD -0.66(-1.18 to -0.14)		Very low	7
Female infertility[l] (an SMD positive value indicates a detrimental RF-EMF effect)													
4	CES***	-1	0	0	-1	0	0	69	83	SMD 0.08(-0.39 to 0.55)		Low	7

Explanations of grading.

[a] Litter size. Risk of Bias: all but two relevant studies were of "some concern".

[b] Resorbed or dead fetuses. Risk of Bias: all relevant studies were of "some concern". Inconsistency: wide range of individual study OR values, heterogeneity only partly explained by subgroup analysis for SAR. Consistency across species: not upgraded because the heterogeneity in the species subgroups is borderline significant ($p = 0.07$).

[c] Fetal weight. Risk of Bias: all relevant studies but one were of "some concern". Inconsistency: because I^2 is 73% and only subgrouping by SAR explains partially the heterogeneity of individual studies.

[d] Fetal length. Risk of Bias: all relevant studies were of "some concern". Inconsistency: because I^2 is 77% and only subgrouping by SAR and temperature increase explains partially the heterogeneity of individual studies. Consistency among species: not upgraded because the database is limited to 13 studies of which only 3 on mice.

[e] Fetal malformations. Risk of Bias: all relevant studies but one were of "some concern". Publication bias: upon visual inspection and Egger test $p = 0.07$ publication bias is assumed by expert judgement.

[f] Litters with malformed fetuses. Risk of Bias: all relevant studies were of "some concern". Inconsistency: extremely wide range of individual study OR values. Indirectness: litter with malformed fetuses is a rough estimate of malformation incidence. Publication bias: upon visual inspection and Egger test $p = 0.09$ publication bias is assumed by expert judgement.

[g] Sex ratio. Risk of Bias: all relevant studies but one were of "some concern". Indirectness: the endpoint is a very weak indicator of adverse pregnancy, not supported by a likely adverse outcome pathway. Consistency among species: not upgraded because the database is limited to 13 studies of which only 2 on mice.

[h] Brain pathology. Risk of Bias: all relevant studies were of "some concern".

[i] Learning and memory functions. Risk of Bias: all relevant studies were of "some concern". Imprecision: CI boundaries cross the null hypothesis and the experimental group sizes are very small. Publication bias: upon visual inspection and Egger test p less than 0.001 publication bias is assumed.

[j] Motor activity functions. Risk of Bias: all relevant studies were of "some concern". Inconsistency: 4 studies conducted at the same exposure level under the same exposure conditions 2 in males and 2 in females, all on the same side of the null hypothesis but quite variable. Imprecision: CI boundaries do not cross the null hypothesis, but the group sizes are small, barely reaching the optimal information size.

[k] Motor and sensory functions. Risk of Bias: all relevant studies were of "some concern". Imprecision: CI boundaries do not cross the null hypothesis, but the group sizes are small and the number of studies is extremely limited.

[l] Female infertility. Risk of Bias: all relevant studies were of "some concern". Imprecision: CI boundaries cross the null hypothesis with a wide confidence interval.

[*] all studies considered irrespective of their RoB rating.

[**] the importance of each database in relation to human possible adverse pregnancy outcomes was rated on a scale 1–10 from the least to the most important.

[***] Controlled Experimental Studies.

for the final assessment of the body of evidence.

3.6. Reporting bias assessment

Reporting bias was indicated by borderline significant results of the Egger's test in the case of malformations as continuous or binary variables and by visual inspection of the funnel plot and Egger's test results in the case of learning and memory functions (Supplementary File 10). There was no evidence of reporting bias for any of the other endpoints.

4. Discussion

4.1. Summary of the evidence and interpretation of the results

Findings have been evaluated according to a GRADE approach as shown in Table 5.

4.1.1. Reduction of fecundity

For pre-implantation loss, only one study was retrieved at "low or some concern" RoB level and, for this reason, this result was not assessed by the GRADE approach. There was high certainty evidence that RF-EMF does not influence litter size (MD 0.05 pups, 95% CI −0.21 to 0.3, 24 studies). There was low certainty evidence that RF-EMF may increase the incidence of resorbed or dead fetuses (OR 1.84, 95% CI 1.27 to 2.66, 21 studies). The inconsistency between the overall results of the *meta*-analyses on litter size and dead fetuses data can be likely explained by the ~ 5 times higher whole body average SAR and twice as many studies conducted at or above 5 W/kg in the latter set of data.

The possibility that high RF-EMF exposure levels may reduce fecundity is reinforced by the analysis of studies reporting a temperature increase in the exposed dams higher than 1 °C. Litter size was reduced only in the 4 studies measuring an average dam temperature increase of 2 °C (MD 0.99, 95% CI 0.36 to 1.62) and the incidence of resorbed/dead fetuses was increased only in the 14 studies measuring an average dam temperature increase of 3 °C (OR 1.59, 95% CI 1.15 to 2.19). The dose–response relationship for the incidence of resorbed/dead fetuses supports the results of the *meta*-analysis with a small but significant linear OR increase with increasing RF-EMF exposure level.

4.1.2. Adverse effects on the offspring health at birth

The *meta*-analysis on fetal weight shows a small but significant decrease of weight in the exposed offspring (SMD 0.31, 95% CI 0.15 to 0.48, 48 studies) at a whole body average SAR of 9.83 W/kg. A moderate level of certainty was attributed to this evidence. Results on fetal weight are consistent with the *meta*-analysis on fetal length showing a moderate detrimental effect on this endpoint (SMD 0.45, 95% CI 0.07 to 0.83, 13 studies) at a whole body average SAR exposure level of 4.55 W/kg, to which a low certainty of evidence was attributed. We synthesized studies on the incidence of malformations at birth in two separate *meta*-analyses according to their reporting data as continuous or binary variables. The *meta*-analysis of continuous data showed a moderate detrimental effect of RF-EMF exposure (SMD −0.45, 95% CI −0.68 to −0.23, 13 studies) confirmed by the statistically significant OR calculated for the binary data (OR 3.22, 95% CI 1.90 to 5.46, 28 studies). We rated the certainty of evidence of continuous data as low and that of binary data as very low. RF-EMF exposure may not affect sex ratio as shown by the pooled effect size of the *meta*-analysis (OR 1.08, 95% CI 0.92 to 1.28, 13 studies). We attributed low certainty to this result.

The RF-EMF effect on fetal weight reduction appears to be due to the contribution of studies that tested whole body exposure SARs equal to or higher than 5 W/kg (SMD 0.51, 95% CI 0.31 to 0.72, 26 studies). Similarly, the contribution of studies at the highest tested SARs (SMD 0.92, 95% CI 0.47 to 1.37, 5 studies) seems to account for the observed RF-EMF effect on fetal length. These observations are supported by the analysis of the dose–response relationships that showed a dose related impact of exposure on both endpoints. However, the possibility that RF-

EMF exposure might be associated to a developmental delay deserves further investigation as the analysis of studies on fetal length by dam core temperature increase shows a temperature related effect, while in the case of studies on fetal weight the effects do not correlate with dam temperature increase (Supplementary File 8).

The studies on malformations as continuous or binary variables were conducted at whole-body average SARs of 6.75 and 16.63 W/kg, respectively. Only the pooled effect sizes of the studies conducted at a SAR equal to or higher than 5 W/kg showed a significant effect (SMD −0.66, 95% CI −0.84 to −0.48, 9 studies and OR 3.49, 95% CI 1.91 to 6.39, 24 studies, for continuous and binary data, respectively), suggesting a dose-related detrimental impact of RF-EMF exposure. These observations are supported by the analysis of the dose–response relationships. The analysis of studies on malformations by dam core temperature increase supports the evidence of an effect at high exposure levels, as shown by the statistically significant pooled effect sizes of the subgroups in which a temperature increase higher that 1 °C was measured (SMD −0.68, 95% CI −0.87 to −0.49, 7 studies and OR 4.53, 95% CI 2.08 to 9.85, 19 studies, for continuous and binary data, respectively).

The analysis of studies on sex ratio by SAR and dam core temperature increase shows that neither variable has a differential impact on this endpoint.

4.1.3. Delayed effects on the offspring health

RF-EMF exposure during pregnancy probably does not have an effect on the offspring brain or cerebellum weight, as shown by the pooled effect size of the *meta*-analysis (SMD 0.10, 95% CI −0.09 to 0.29, 12 studies). We attributed a moderate certainty to this result. RF-EMF exposure during pregnancy may have a moderate detrimental effect on learning and memory functions as measured by the maze escape latency time of the offspring (SMD −0.54, 95% CI −1.24 to 0.17, 2 studies). However, we attributed a very low certainty to this result. RF-EMF exposure during pregnancy may have a large detrimental effect on motor activity functions as measured by the endurance time of the offspring in any type of test (SMD 0.79, 95% CI 0.21 to 1.38, 4 studies) but, again, we are uncertain of the result. RF-EMF exposure during pregnancy may have a moderate to large detrimental effect on motor and sensory functions as measured by the magnitude of the startle response of the offspring (SMD −0.66, 95% CI −1.18 to −0.14, 2 studies), but we are uncertain of this result. For behavioural ontogeny no study was retrieved at "low or some concern" RoB level and, for this reason, this endpoint was not assessed by the GRADE approach. RF-EMF exposure during pregnancy may not affect the F2 litter size (SMD 0.08, 95% CI −0.39 to 0.55, 4 studies). A low certainty was attributed to this result.

In conclusion, studies on experimental mammals indicate that RF-EMF exposure does not have a detrimental effect on fecundity based on the high level of certainty for results on litter size. There is a moderate certainty that RF-EMF exposure likely affect offspring at birth, based on the *meta*-analysis of studies on fetal weight. There is a moderate certainty that RF-EMF exposure does not have a delayed effect on the weight of brain or cerebellum after *in utero* exposure. On the other hand, RF-EMF may have a delayed adverse effect, varying in magnitude on neurobehavioural functions, but these findings are very uncertain. Finally, our results show that RF-EMF exposure of experimental mammals *in utero* may not have a delayed effect on the fertility of the female offspring.

4.2. Limitations in the evidence

Although the database of relevant studies was not small, including 88 papers, 2 main reasons limited the body of evidence suited for a *meta*-analysis or at least a synthetic overview. One reason was the heterogeneity of the endpoints employed to evaluate RF-EMF effects on pregnancy, which reduced the number of studies for each *meta*-analysis. A second reason was the quality of the studies. Many studies were scored at an overall "some concern" RoB mainly because of poor exposure characterization, lack of blinding during performance of the experiments or outcome assessment and/or suboptimal animal sample size. In some cases, these same reasons led to a "high concern" RoB rating that entailed the exclusion of their results from the body of evidence on which conclusions were drawn and weighed according to the GRADE approach. In general, the quality of studies seemed to reflect the research investment, as shown by the association of poor confidence in the exposure set-up and dosimetry ("definitely high RoB" for question about confidence in exposure characterization) with smaller experimental group sizes.

A large heterogeneity of study characteristics also posed difficulties for the review. The evolution of research goals in the literature, from an initial interest about RF-EMF thermal effects to the more recent concern about mobile phone low level RF-EMF emissions, probably contributed to this problem.

Although 3 studies included animals exposed to direct heating in addition to sham exposed and RF-EMF exposed animals (Chernovetz et al., 1977; Nawrot et al., 1981, 1985), in no study these animals strictly matched to RF-EMF exposed ones for core body temperature increase and could be used as a temperature comparator to specifically evaluate heat-independent RF-EMF effects. In hindsight, we think this to be very difficult to achieve in animal studies and we would not recommend investing resources in experiments aimed at this goal.

A specific limitation of the studies on neurobehavioural effects was the small number of those amenable for the *meta*-analysis which was worsened by the lack of independent replication, since in many cases they were derived from only a few laboratories.

No study was retrieved reporting results on ano-genital distance at birth or early onset cancer after RF-EMF *in utero* only exposure. This is a limitation of the literature database that should possibly be addressed in future research.

In addition, most papers did not follow a standardized experimental design and reporting of results, as recommended by international guidelines (OECD TG 414, OECD TG 426) likely because the intent of most studies was exploratory research rather than risk assessment. Major issues consisted of considering the single pup instead of the litter as the experimental unit or not clarifying which one was the case, and matching a single sham-exposed comparator group to different exposed groups. Finally, the variety of parameters and metrics used to measure the same endpoint made it difficult to conduct a *meta*-analysis and forced us to make choices as detailed in Supplementary File 2.

Finally, publication bias was suspected for evidence related to fetal malformations and learning and memory functions.

4.3. Limitations in the review process

We could not make a decision about the inclusion of 21 papers out of the 236 selected by title/abstract examination because we could not retrieve 11 papers and were unable to translate 10 papers. However, given the large number of included studies, we doubt that this would have influenced our conclusions.

We decided to base the *meta*-analysis of delayed pathological effects only on the offspring brain upon weight measurement, because we considered the count of the cell number in histopathological sections a less standardized method. We reviewed the latter data only in a narrative way. Similarly, for the *meta*-analysis of data on learning and memory functions in the offspring, we chose to use only maze escape latency time, and extracted data only for the most challenging tests. Data relative to other markers were presented in a narrative way. Finally, for the *meta*-analysis of data assessing the offspring motor activity performance, we used only the time of endurance in any of the applied tests and narrated papers reporting data using other metrics.

For the inclusion in the *meta*-analysis of studies in which a single comparator was matched to different exposed groups of animals, we had to make a choice between multiple options while avoiding including

study participants more than once in the analysis: arbitrarily choosing one exposure level to match the single comparator, dividing the number of animals of the control group by the number of the exposed groups or renouncing to the independency of results among the exposed groups by averaging their exposure conditions and responses. We chose the last option because, in our opinion, it introduced less bias and we had decided to explore the dose–effect relationship through dose–effect *meta*-analysis.

For the studies that ignored the clustering of effects among pups in one litter, the so-called litter effect, and that used the individual pups as the experimental unit, we made adjustments to the reported effects. We either used a conservative approach referring the mean values and variation parameters to the number of dams, as in the case of continuous data, or applied an intra-cluster correction factor of 0.2, as in the case of binary data. For delayed effects, we could only use pups as the experimental unit because the number of dams was not reported in many studies. This will have overestimated the precision of the effects in these studies.

We acknowledge limitations in our subgroup analyses. In particular, the choice of conducting subgroup analysis to investigate sources of heterogeneity among the studies even when the subgroups were small (N = 3 studies) induces the risk of false negative outcomes. However, for the interpretation of results we relied upon the statistical significance of the between-group difference, which takes into account the group size. In addition, we acknowledge the limitation of subgrouping the studies showing an increase of dam core temperature below or above 1 °C, since it does not take into account the magnitude, timing or duration of temperature increase. Nevertheless, for some of the endpoints, this analysis confirmed the hypothesis of the adverse effects of maternal heating during pregnancy. Similarly, the choice of subgrouping the studies by exposure level in only 3 groups, due to the need to include a minimum of 3 studies in each group for the majority of endpoints, may have blurred the contribution of studies testing very high exposure levels. Nevertheless, the results of the subgroup analysis were consistent with the hypothesis of larger effects at SAR levels above 5 W/kg.

For assessing publication bias, we visually inspected the funnel plots of studies followed by statistical evaluation by Egger's test. We acknowledge the limits of the Egger's test and the possibility of false positive results when drawing funnel plots with large SMD effect sizes (Zwetsloot et al., 2017), as in the case of fetal malformations and learning and memory functions. However, considering that no better alternatives exist to the Egger's test, that the certainty of evidence attributed to the results on learning and memory functions would not change, and that the certainty of evidence to the results on fetal malformations would change from "very low" to "low", we do not think this is a major limitation of the systematic review.

Finally, it cannot be excluded that the pooled effect sizes resulting from our *meta*-analyses were exaggerated in the direction of an RF-EMF effect. The comparison of pooled effect sizes between "low or some concern" and "high concern" studies (Table 4) shows a higher effect for some of the most relevant endpoints in the "high concern" studies. Unfortunately, so few studies could be classified as "low concern" that such a comparison could not be done between "low" vs "some concern" and we were forced to include all these studies in the one category "low or some concern" studies. Nevertheless, the trend observed in Table 4 suggests that the evidence of an RF-EMF effect on pregnancy and birth outcomes could be overstated to a certain degree.

In spite of these limitations, we believe that our results advance the state-of-the-art knowledge for a comprehensive assessment of the body of evidence available in peer reviewed literature. We aimed at inclusiveness, not posing limitations to language and year of publication, and considered the many facets of adverse pregnancy outcomes. We applied an internationally standardized methodology for systematic review and *meta*-analysis, including the OHAT method for risk of bias assessment (NTP 2015a, b).

4.4. Implications for policy and research

This systematic review of animal studies shows that RF-EMF exposure does not affect fecundity and likely has only a small effect on fetal weight decrease. However, some studies retrieved by the literature search that showed a detrimental effect on the incidence of dead/resorbed fetuses or the increase of malformations at high exposure levels, largely exceeding the current human exposure limits, cannot be discounted. These studies confirm what is known about the harmful effect of heating on fetuses, but they leave largely uncertain the possibility of RF-EMF effects at lower exposure levels, closer to relevant human exposure levels. Currently, it remains difficult to determine the exposure levels at which RF-EMF can start to affect fecundity or offspring health at birth. The whole body average SAR values in the included experiments are well above the recommended human exposure limit values for the general public set by international bodies (ICNIRP 2020). Actual SAR values experienced by the public in the general environment are below, and in most cases, well below, the recommended human exposure limit values. The dose effect *meta*-analyses contributed to support the results of the *meta*-analysis but were not supposed to define the shape of the dose–effect relationship or find a minimum exposure level at which a clear effect could be discerned.

For two endpoints planned in the protocol, namely ano-genital distance at birth and early-onset cancer no studies were retrieved. Anogenital distance is a well-known developmental biomarker associated with impairment of the reproductive system and exposure to environmental carcinogens during pregnancy has been linked with development of childhood cancer (Botsivali and Kyrtopoulos, 2019). Hopefully, future research will shed light on the impact of RF-EMF exposure on these outcomes.

As a whole, the possible impact of *in utero* RF-EMF exposure remains uncertain due to the severe limitations of some of the studies. In particular, during the systematic review, we identified several methodological limitations in the studies that should be overcome in future studies to improve the quality of the research. Blinding during experiment performance and outcome assessment should always be applied to minimize bias. More adherence to OECD Test Guideline 414 "Prenatal Developmental Toxicity Study" and 426 "Developmental Neurotoxicity Study" is recommended together with a more standardized approach for reporting results. A large proportion of included studies was rated at either "some" or "high concern" for RoB for exposure characterisation or temperature rise assessment and some others had to be excluded from the systematic review because they did not reach a minimum quality standard for these aspects. We would recommend that future studies bear the reasons for exclusion or RoB concerns in mind in study design and implementation. There are several papers in the research literature with recommendations on how exposure characterisation concerns can be mitigated, for example Kuster and Schönborn (2000). Finally, studies investigating not just a single level but several exposure levels, spanning from low levels comparable to human exposure to higher levels where mild hyperthermic effects could be expected, should be conducted under the same experimental conditions.

In spite of the large number of studies collected, our systematic review could only partly answer the PECO question and did not provide conclusions certain enough to inform decisions at a regulatory level, but it can be considered a solid starting point to direct future research on this topic.

4.5. Registration and protocol

4.5.1. Protocol registration

The protocol for the systematic review was published in Pacchierotti et al., 2021 and registered in PROSPERO (CRD42021227746, https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=227746).

4.5.2. Deviations from the protocol

1. Minor changes consisted of a different organization and a slight rewording of the outcomes and endpoints as reported in Supplementary File 1, in which specific reasons for changes are also shown. These changes led to a slight rewording of the PECO question, as shown in the same file.
2. Some specific decisions on which data to extract were taken after the publication of the protocol but before inspecting the results (see Supplementary File 2).
3. To base the assessment of possible RF-EMF exposure impact on the most solid set of data, we excluded studies at "high concern" for RoB from the summary of findings assessed for the certainty of evidence by the GRADE approach, even if this had not been explicitly stated in the protocol.
4. In relation to exposure eligibility criteria, the protocol specified that studies in which exposure level from mobile phones or other RF-EMF generating devices was not measured or estimated by reliable methods, but simply inferred from assumed exposure conditions from the RF-generating device type, were to be assessed as a separate group. We preferred to assess this group together with the other studies and rate the confidence in the exposure characterization by the RoB assessment, because it was difficult to set boundaries in a continuum of exposure dosimetry quality reporting. In addition, we specified the exclusion of studies on exposure to ultrasound.
5. Among the reasons for exclusion, we added papers not retrieved or not translated that had not been foreseen at the protocol stage.
6. For binary outcomes, we used Odds Ratio instead of Relative Risk as the effect size measure because it was more easily tractable by the applied data analysis software.
7. Information regarding conflict-of-interest declarations and funding sources were not analysed since public funding and absence of conflict of interest were declared in the vast majority of papers.
8. Among the factors envisaged in the Protocol, we limited our heterogeneity analyses to exposure levels and dam core temperature increase, because these are the variables most likely affecting RF-EMF biological effects, and experimental animal species, because inter-species consistency of results was to be considered as an upgrading factor for the certainty of evidence. We did not explore sources of heterogeneity by differences in tested radiofrequencies because only 3 endpoints (with no more than 3 studies each) were assessed at frequencies above 6000 MHz. This was the upper range in which a different mechanism of biological interaction might be expected because of short penetration depth into superficial tissues (a few mm or less).

CRediT authorship contribution statement

Eugenia Cordelli: Conceptualization, Data curation, Funding acquisition, Investigation, Project administration, Supervision, Writing - original draft, Writing - review & editing. **Lucia Ardoino:** . **Barbara Benassi:** . **Claudia Consales:** . **Patrizia Eleuteri:** . **Carmela Marino:** . **Maurizio Sciortino:** . **Paola Villani:** . **Martin H. Brinkworth:** . **Guangdi Chen:** . **James P. McNamee:** . **Andrew W. Wood:** . **Lea Belackova:** . **Jos Verbeek:** . **Francesca Pacchierotti:** .

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: AWW previously directed a research group, which included two technical associates who are telecommunications company employees. AWW has been member of the ICNIRP Scientific Expert Group (SEG) from 2013 until 2021 and collaborates with the Australian Radiation Protection and Nuclear Safety Agency. JPMN was a member for IARC Monograph 102 Working Group assessing the carcinogenicity of RF-EMF (Mechanistic Studies sub-group), a co-author of Canada's Safety Code 6 (which are the *de facto* national human exposure limits applied in Canada) and a member of the WHO EMF Project International Advisory Committee (Canadian representative). Health Canada financially contributed to the WHO EMF Project to support the completion of the systematic reviews on RF-EMF. CM has been member of Technical Consultation on the WHO RF Research Agenda (2010), member of ICNIRP main commission since May 2012, confirmed in 2016 and 2020, Italian delegate for the European Cost Actions BM0704 and BM1309 "EMF-MED". All other authors declare that they have no known conflicts of interest.

Data availability

Individual study results included in the systematic review and *meta*-analyses are reported in Supplementary File 7. Extraction forms are available upon request

Acknowledgements

We are grateful to Emilie van Deventer, Maria Rosaria Scarfì and Eric van Rongen for advice regarding the protocol draft and for discussions to ensure consistency in approaches across the multiple ongoing WHO systematic reviews. We wish to thank Flavio Di Marzio for his appreciated graphical help.

Funding

This project was partially funded by the World Health Organization (contracts 2020/1026306-0, 2022/1275453-1). WHO provided the basis for the protocol and methodological support throughout the review process. Additional in-kind funds were provided by ENEA, Health Canada and Swinburne University of Technology.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.envint.2023.108178.

References

Advisory Group on Non-ionising Radiation (AGNIR), 2012. Health Effects from Radiofrequency Electromagnetic Fields. Report of the independent Advisory Group on Non-ionising Radiation. UK Health Protection Agency.

Aït-Aïssa, S., Billaudel, B., Poulletier de Gannes, F., Ruffié, G., Duleu, S., Hurtier, A., Haro, E., Taxile, M., Athané, A., Geffard, M., Wu, T., Wiart, J., Bodet, D., Veyret, B., Lagroye, I., 2012. In utero and early-life exposure of rats to a Wi-Fi signal: screening of immune markers in sera and gestational outcome. Bioelectromagnetics 33, 410–420. https://doi.org/10.1002/bem.21699.

Albert, E.N., Sherif, M.F., Papadopoulos, N.J., Slaby, F.J., Monahan, J., 1981. Effect of nonionizing radiation on the purkinje cells of the rat cerebellum. Bioelectromagnetics 2, 247–257. https://doi.org/10.1002/bem.2250020306.

Alchalabi, A.S.H., Aklilu, E., Aziz, A.R., Malek, F., Ronald, S.H., Khan, M.A., 2016. Different periods of intrauterine exposure to electromagnetic field: Influence on female rats' fertility, prenatal and postnatal development. Asian Pacific Journal of Reproduction 5, 14–23. https://doi.org/10.1016/j.apjr.2015.12.003.

Alchalabi, A.S., Aklilu, E., Aziz, A., Rahim, H., Ronald, S., Malek, M., Khan, M., 2017. Impact of electromagnetic radiation exposure during pregnancy on embryonic skeletal development in rats. Asian Pacific Journal of Reproduction 6, 104–111. https://doi.org/10.12980/apjr.6.20170302.

Aldad, T.S., Gan, G., Gao, X.B., Taylor, H.S., 2012. Fetal radiofrequency radiation exposure from 800–1900 mhz-rated cellular telephones affects neurodevelopment and behavior in mice. Sci. Rep. 2, 312. https://doi.org/10.1038/srep00312.

Anderson, L.E., Sheen, D.M., Wilson, B.W., Grumbein, S.L., Creim, J.A., Sasser, L.B., 2004. Two-year chronic bioassay study of rats exposed to a 1.6 GHz radiofrequency signal. Radiat. Res. 162, 201–210. https://doi.org/10.1667/rr3208.

Azimzadeh, M., Jelodar, G., 2020. Prenatal and early postnatal exposure to radiofrequency waves (900 MHz) adversely affects passive avoidance learning and memory. Toxicol. Ind. Health 36, 1024–1030. https://doi.org/10.1177/0748233720973143.

Baldacci, S., Gorini, F., Santoro, M., Pierini, A., Minichilli, F., Bianchi, F., 2018. Environmental and individual exposure and the risk of congenital anomalies: a

review of recent epidemiological evidence. Epidemiol. Prev. 42, 1–34. https://doi.org/10.19191/ep18.3-4.S1.P001.057.
Bas, O., Sönmez, O.F., Aslan, A., Ikinci, A., Hanci, H., Yildirim, M., Kaya, H., Akca, M., Odaci, E., 2013. Pyramidal Cell Loss in the Cornu Ammonis of 32-day-old Female Rats Following Exposure to a 900 Megahertz Electromagnetic Field During Prenatal Days 13–21. Neuroquantology 11, 591–599. https://doi.org/10.14704/nq.2013.11.4.701.
Berman, E., Carter, H.B., 1984. Decreased body weight in fetal rats after irradiation with 2450-MHz (CW) microwaves. Health Phys. 46, 537–542. https://doi.org/10.1097/00004032-198403000-00004.
Berman, E., Kinn, J.B., Carter, H.B., 1978. Observations of mouse fetuses after irradiation with 2.45 GHz microwaves. Health Phys. 35, 791–801. https://doi.org/10.1097/00004032-197812000-00006.
Berman, E., Carter, H.B., House, D., 1981. Observations of rat fetuses after irradiation with 2450-MHz (CW) microwaves. J. Microw. Power 16, 9–13. https://doi.org/10.1080/16070658.1981.11689226.
Berman, E., Carter, H.B., House, D., 1982a. Reduced weight in mice offspring after in utero exposure to 2450-MHz (CW) microwaves. Bioelectromagnetics 3, 285–291. https://doi.org/10.1002/bem.2250030212.
Berman, E., Carter, H.B., House, D., 1982b. Observations of Syrian hamster fetuses after exposure to 2450-MHz microwaves. J. Microw. Power 17, 107–112. https://doi.org/10.1080/16070658.1982.11689270.
Berman, E., Carter, H.B., House, D., 1984. Growth and development of mice offspring after irradiation in utero with 2,450-MHz microwaves. Teratology 30, 393–402. https://doi.org/10.1002/tera.1420300312.
Berman, E., Weil, C., Philips, P.A., Carter, H.B., House, D.E., 1992. Fetal and maternal effects of continual exposure of rats to 970-mhz circularly-polarized microwaves. Electromagn. Biol. Med. 11, 43–54. https://doi.org/10.3109/15368379209012851.
Blencowe, H.; Krasevec, J.; de Onis, M.; Black, R.E.; An, X.; Stevens, G.A.; Borghi, E.; Hayashi, C.; Estevez, D.; Cegolon, L.; Shiekh, S.; Ponce Hardy, V.; Lawn, J.E.; Cousens, S. 2019; National, regional, and worldwide estimates of low birthweight in 2015, with trends from 2000: a systematic analysis. Lancet Glob Health 7: e849-e860. doi:10.1016/s2214-109x(18)30565-5.
Bornhausen, M., Scheingraber, H., 2000. Prenatal exposure to 900 MHz, cell-phone electromagnetic fields had no effect on operant-behavior performances of adult rats. Bioelectromagnetics 21, 566–574.
Botsivali, M., Kyrtopoulos, S.A., 2019. Transplacental exposure to carcinogens and risks to children: evidence from biomarker studies and the utility of omic profiling. Arch. Toxicol. 93, 833–857. https://doi.org/10.1007/s00204-019-02428-3.
Brown-Woodman, P.D.C., Hadley, J.A., 1988a. Teratogenic Effects of Exposure to Radiofrequency Radiation (27.12 MHz) from a Shortwave Diathermy Unit. Ind. Health 26, 1–10. https://doi.org/10.2486/indhealth.26.1.
Brown-Woodman, P.D.C., Hadley, J.A., 1988b. Studies of the Teratogenic Potential of Exposure of Rats to 27.12 MHZ Pulsed Shortwave Radiation. Journal of Bioelectricity 7, 57–67. https://doi.org/10.3109/15368378809016177.
Calis, P., Seymen, M., Soykan, Y., Delen, K., Aral, B.S., Take Kaplanoglu, G., Karcaaltincaba, D., 2019. Does Exposure of Smart Phones during Pregnancy Affect the Offspring's Ovarian Reserve? A Rat Model Study. Fetal Pediatr. Pathol. 1–11. https://doi.org/10.1080/15513815.2019.1692112.
Chazan, B., Janiak, M., Kobus, M., Marcickiewicz, J., Troszyński, M., Szmigielski, S., 1983. Effects of microwave exposure in utero on embryonal, fetal and postnatal development of mice. Biol. Neonate 44, 339–348. https://doi.org/10.1159/000241749.
Chernovetz, M.E., Justesen, D.R., King, N.W., Wagner, J.E., 1975. Teratology, survival, and reversal learning after fetal irradiation of mice by 2450-MHz microwave energy. J. Microw. Power 10, 391–409.
Chernovetz, M.E., Justesen, D.R., Oke, A.F., 1977. A teratological study of the rat: Microwave and infrared radiations compared. Radio Sci. 12, 191–197. https://doi.org/10.1029/RS012i06Sp00191.
Chiang, H., 1988. Effects of pre- and post-natal pulse microwave irradiation on developing mice. Chin. J. Biomed. Eng. 7, 63–69.
Cobb, B.L., Jauchem, J.R., Mason, P.A., Dooley, M.P., Miller, S.A., Ziriax, J.M., Murphy, M.R., 2000. Neural and behavioral teratological evaluation of rats exposed to ultra-wideband electromagnetic fields. Bioelectromagnetics 21, 524–537. https://doi.org/10.1002/1521-186x(200010)21:7<524::aid-bem6>3.0.co;2-j.
Dasdag, S., Akdag, Z., Ayyildiz, O., Demirtas, O.C., Yayla, M., Sert, C., 2000. Do cellular phones alter blood parameters and birth weight of rats? Electro Magnetobiol 19, 107–113.
DastAmooz, S., Tahmasebi Boroujeni, S., Shahbazi, M., Vali, Y., 2018. Physical activity as an option to reduce adverse effect of EMF exposure during pregnancy. Int. J. Dev. Neurosci. 71, 10–17. https://doi.org/10.1016/j.ijdevneu.2018.07.009.
Egger, M., Davey Smith, G., Schneider, M., Minder, C., 1997. Bias in meta-analysis detected by a simple, graphical test. BMJ 315, 629–634. https://doi.org/10.1136/bmj.315.7109.629.
Erdem Koç, G., Kaplan, S., Altun, G., Gümüş, H., Gülsüm Deniz, Ö., Aydin, I., Emin Onger, M., Altunkaynak, Z., 2016. Neuroprotective effects of melatonin and omega-3 on hippocampal cells prenatally exposed to 900 MHz electromagnetic fields. Int. J. Radiat Biol. 92, 590–595. https://doi.org/10.1080/09553002.2016.1206223.
Ferreira, A.R., Knakievicz, T., Pasquali, M.A., Gelain, D.P., Dal-Pizzol, F., Fernández, C. E., de Salles, A.A., Ferreira, H.B., Moreira, J.C., 2006. Ultra high frequency-electromagnetic field irradiation during pregnancy leads to an increase in erythrocytes micronuclei incidence in rat offspring. Life Sci. 80, 43–50. https://doi.org/10.1016/j.lfs.2006.08.018.
Galvin, M.J., MacNichols, G., McRee, D.I., 1983. Effects of 2450 MHz microwave radiation during the gestational period on the postnatal hematology of rats. Cell Biophys. 5, 33–41. https://doi.org/10.1007/bf02788585.
Galvin, M.J., Tilson, H.A., Mitchell, C.L., Peterson, J., McRee, D.I., 1986. Influence of pre- and postnatal exposure of rats to 2.45-GHz microwave radiation on neurobehavioral function. Bioelectromagnetics 7, 57–71. https://doi.org/10.1002/bem.2250070107.
Golub, M.S., Sobin, C.A., 2020. Statistical modeling with litter as a random effect in mixed models to manage "intralitter likeness". Neurotoxicol. Teratol. 77, 106841 https://doi.org/10.1016/j.ntt.2019.106841.
Guler, G., Tomruk, A., Ozgur, E., Seyhan, N., 2010. The effect of radiofrequency radiation on DNA and lipid damage in non-pregnant and pregnant rabbits and their newborns. Gen. Physiol. Biophys. 29, 59–66. https://doi.org/10.4149/gpb_2010_01_59.
Guyatt, G., Oxman, A.D., Akl, E.A., Kunz, R., Vist, G., Brozek, J., Norris, S., Falck-Ytter, Y., Glasziou, P., DeBeer, H., Jaeschke, R., Rind, D., Meerpohl, J., Dahm, P., Schünemann, H.J., 2011. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. J. Clin. Epidemiol. 64, 383–394. https://doi.org/10.1016/j.jclinepi.2010.04.026.
Haghani, M., Shabani, M., Moazzami, K., 2013. Maternal mobile phone exposure adversely affects the electrophysiological properties of Purkinje neurons in rat offspring. Neuroscience 250, 588–598. https://doi.org/10.1016/j.neuroscience.2013.07.049.
Heynick, L.N.; Merritt, J.H. 2003; Radiofrequency fields and teratogenesis. Bioelectromagnetics Suppl 6: S174-186. doi:10.1002/bem.10127.
Higgins JPT, Li T, Deeks JJ (editors). 2022. Chapter 6: Choosing effect measures and computing estimates of effect. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.3 (updated February 2022). Cochrane. Available from www.training.cochrane.org/handbook.
Hooijmans, C.R., Rovers, M.M., de Vries, R.B., Leenaars, M., Ritskes-Hoitinga, M., Langendam, M.W., 2014. SYRCLE's risk of bias tool for animal studies. BMC Med. Res. Method. 14, 43. https://doi.org/10.1186/1471-2288-14-43.
Hooijmans, C.R., de Vries, R.B.M., Ritskes-Hoitinga, M., Rovers, M.M., Leeflang, M.M., IntHout, J., Wever, K.E., Hooft, L., de Beer, H., Kuijpers, T., Macleod, M.R., Sena, E. S., Ter Riet, G., Morgan, R.L., Thayer, K.A., Rooney, A.A., Guyatt, G.H., Schünemann, H.J., Langendam, M.W., 2018. Facilitating healthcare decisions by assessing the certainty in the evidence from preclinical animal studies. PLoS One 13, e0187271.
Ikinci, A., Odaci, E., Yildirim, M., Kaya, H., Akça, M., Hanci, H., Aslan, A., Sönmez, O.F., Baş, O., 2013. The effects of prenatal exposure to a 900 megahertz electromagnetic field on hippocampus morphology and learning behavior in rat pups. NeuroQuantology 11, 582–590. https://doi.org/10.14704/nq.2013.11.4.699.
Inaloz, S.S., Dasdag, S., Aslan, A., Bilici, A., Yayla, M., 1997. The effects of the microwave oven on pregnant and newborn rat kidneys. Urogynaecologia International Journal 11, 9–15.
Inouye, M., Galvin, M.J., McRee, D.I., 1983. Effect of 2,450 MHz microwave radiation on the development of the rat brain. Teratology 28, 413–419. https://doi.org/10.1002/tera.1420280312.
International Commission on Non-Ionizing Radiation Protection (ICNIRP), 2020. Guidelines for Limiting Exposure to Electromagnetic Fields (100 kHz to 300 GHz). Health Phys. 118, 483–524.
Jensh, R.P., 1984a. Studies of the teratogenic potential of exposure of rats to 6000-MHz microwave radiation. II. Postnatal psychophysiologic evaluations. Radiat. Res. 97, 282–301.
Jensh, R.P., 1984b. Studies of the teratogenic potential of exposure of rats to 6000-MHz microwave radiation. I. Morphologic analysis at term. Radiat. Res. 97, 272–281.
Jensh, R.P., Vogel, W.H., Brent, R.L., 1982a. Postnatal functional analysis of prenatal exposure of rats to 915 MHz microwave radiation. J. Am. Coll. Toxicol. 1, 73–90. https://doi.org/10.3109/10915818209018019.
Jensh, R.P., Weinberg, I., Brent, R.L., 1982b. Teratologic studies of prenatal exposure of rats to 915-MHz microwave radiation. Radiat. Res. 92, 160–171. https://doi.org/10.2307/3575851.
Jensh, R.P., Weinberg, I., Brent, R.L., 1983a. An evaluation of the teratogenic potential of protracted exposure of pregnant rats to 2450-MHz microwave radiation: I. Morphologic analysis at term. J. Toxicol. Environ. Health 11, 23–35. https://doi.org/10.1080/15287398309530318.
Jensh, R.P., Vogel, W.H., Brent, R.L., 1983b. An evaluation of the teratogenic potential of protracted exposure of pregnant rats to 2450-MHz microwave radiation. II. Postnatal psychophysiologic analysis. J. Toxicol. Environ. Health 11, 37–59. https://doi.org/10.1080/15287398309530319.
Kaplan, J.; Polson, P.; Rebert, C.; Lunan, K.; Gage, M. 1982; Biological and behavioral effects of prenatal and postnatal exposure to 2450-MHz electromagnetic radiation in the squirrel monkey. Radio Science 17: 135S-144S. doi:10.1029/RS017i05Sp0135S.
Keles, A.I., Sut, B.B., 2021. Histopathological and epigenetic alterations in the spinal cord due to prenatal electromagnetic field exposure: An H3K27me3-related mechanism. Toxicol. Ind. Health 748233721996947. https://doi.org/10.1177/0748233721996947.
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.
Kubinyi, G.; Thuroczy, G.; Bakos, J.; Boloni, E.; Sinay, H.; Szabo, L.D. 1996; Effect of continuous-wave and amplitude-modulated 2.45 GHz microwave radiation on the liver and brain aminoacyl-transfer RNA synthetases of in utero exposed mice. Bioelectromagnetics 17: 497-503. doi:10.1002/(SICI)1521-186X(1996)17:6<497::AID-BEM10>3.0.CO;2-I.

Kuster, N., Schönborn, F., 2000. Recommended minimal requirements and development guidelines for exposure setups of bio-experiments addressing the health risk concerns of wireless communications. Bioelectromagnetics 21, 508–514. https://doi.org/10.1002/1521-186x(200010)21:7<508::aid-bem4>3.0.co;2-f.

Lary, J.M., Conover, D.L., Foley, E.D., Hanser, P.L., 1982. Teratogenic effects of 27.12 MHz radiofrequency radiation in rats. Teratology 26, 299–309. https://doi.org/10.1002/tera.1420260312.

Lary, J.M., Conover, D.L., Johnson, P.H., 1983. Absence of embryotoxic effects from low-level (nonthermal) exposure of rats to 100 MHz radiofrequency radiation. Scand. J. Work Environ. Health 9, 120–127. https://doi.org/10.5271/sjweh.2434.

Lary, J.M., Conover, D.L., Johnson, P.H., Hornung, R.W., 1986. Dose-response relationship between body temperature and birth defects in radiofrequency-irradiated rats. Bioelectromagnetics 7, 141–149. https://doi.org/10.1002/bem.225007020.

Lary, J.M.; Conover, D.L.; Johnson, P.H.; Burg, J.R. 1983b; Teratogenicity of 27.12-MHz radiation in rats is related to duration of hyperthermic exposure. Bioelectromagnetics 4: 249-255. doi:10.1002/bem.2250040305.

Lee, H.J., Lee, J.S., Pack, J.K., Choi, H.D., Kim, N., Kim, S.H., Lee, Y.S., 2009. Lack of teratogenicity after combined exposure of pregnant mice to CDMA and WCDMA radiofrequency electromagnetic fields. Radiat. Res. 172, 648–652. https://doi.org/10.1667/rr1771.1.

Li, Z.Q., Zhang, Y., Wan, Y.M., Zhou, Q., Liu, C., Wu, H.X., Mu, Y.Z., He, Y.F., Rauniyar, R., Wu, X.N., 2020. Testing of behavioral and cognitive development in rats after prenatal exposure to 1800 and 2400 MHz radiofrequency fields. J. Radiat. Res. 61, 197–206. https://doi.org/10.1093/jrr/rrz097.

Marcickiewicz, J., Chazan, B., Niemiec, T., Sokolska, G., Troszyński, M., Luczak, M., Szmigielski, S., 1986. Microwave radiation enhances teratogenic effect of cytosine arabinoside in mice. Biol. Neonate 50, 75–82. https://doi.org/10.1159/000242571.

Merritt, J.H., Hardy, K.A., Chamness, A.F., 1984. In utero exposure to microwave radiation and rat brain development. Bioelectromagnetics 5, 315–322. https://doi.org/10.1002/bem.2250050304.

National Toxicology Program (NTP), 2015a. Handbook for Conducting a Literature based Health Assessment Using OHAT Approach for Systematic Review and Evidence Integration. Office of Health Assessment and Translation. National Institute of Environmental Health Sciences.

National Toxicology Program (NTP), 2015b. OHAT Risk of Bias Rating Tool for Human and Animal Studies. Office of Health Assessment and Translation.

Nawrot, P.S., McRee, D.I., Staples, R.E., 1981. Effects of 2.45 GHz CW microwave radiation on embryofetal development in mice. Teratology 24, 303–314. https://doi.org/10.1002/tera.1420240309.

Nawrot, P.S., McRee, D.I., Galvin, M.J., 1985. Teratogenic, biochemical, and histological studies with mice prenatally exposed ot 2.45-GHz microwave radiation. Radiat. Res. 102, 35–45. https://doi.org/10.2307/3576427.

Nelson, B.K., Conover, D.L., Brightwell, W.S., Shaw, P.B., Werren, D., Edwards, R.M., Lary, J.M., 1991. Marked increase in the teratogenicity of the combined administration of the industrial solvent 2-methoxyethanol and radiofrequency radiation in rats. Teratology 43, 621–634. https://doi.org/10.1002/tera.1420430618.

Nelson, B.K., Conover, D.L., Shaw, P.B., Werren, D.M., Edwards, R.M., Hoberman, A.M., 1994. Interactive developmental toxicity of radiofrequency radiation and 2-methoxyethanol in rats. Teratology 50, 275–293. https://doi.org/10.1002/tera.1420500403.

Nelson, B.K., Conover, D.L., Krieg Jr, E.F., Snyder, D.L., Edwards, R.M., 1997a. Interactions of Radiofrequency Radiation-Induced Hyperthermia and 2-Methoxyethanol Teratogenicity in Rats. Bioelectromagnetics 18, 349–359. https://doi.org/10.1002/(SICI)1521-186X(1997)18:5<349::AID-BEM2>3.0.CO;2-2.

Nelson, B.K., Conover, D.L., Shaw, P.B., Snyder, D.L., Edwards, R.M., 1997b. Interactions of radiofrequency radiation on 2-methoxyethanol teratogenicity in rats. J. Appl. Toxicol. 17, 31–39. https://doi.org/10.1002/(sici)1099-1263(199701)17:1<31::aid-jat384>3.0.co;2-1.

Nelson, B.K., Snyder, D.L., Shaw, P.B., 1999. Developmental toxicity interactions of salicylic acid and radiofrequency radiation or 2-methoxyethanol in rats. Reprod. Toxicol. 13, 137–145. https://doi.org/10.1016/s0890-6238(98)00071-9.

Nelson, B.K., Snyder, D.L., Shaw, P.B., 2001. Developmental toxicity interactions of methanol and radiofrequency radiation or 2-methoxyethanol in rats. Int. J. Toxicol. 20, 89–100. https://doi.org/10.1080/10915810151115218.

O'Connor, M.E., 1988. Prenatal microwave exposure and behavior. Prog. Clin. Biol. Res. 257, 265–288.

O'Connor, M.E., 1999. Intrauterine effects in animals exposed to radiofrequency and microwave fields. Teratology 59, 287–291. https://doi.org/10.1002/(sici)1096-9926(199904)59:4<287::Aid-tera13>3.0.Co;2-y.

Odaci, E., Ikinci, A., Yildirim, M., Kaya, H., Akça, M., Hanci, H., Sönmez, O.F., Aslan, A., Okuyan, M., Baş, O., 2013. The effects of 900 megahertz electromagnetic field applied in the prenatal period on spinal cord morphology and motor behavior in female rat pups. NeuroQuantology 11, 573–581. https://doi.org/10.14704/nq.2013.11.4.698.

Odaci, E., Hanci, H., Ikinci, A., Sonmez, O.F., Aslan, A., Sahin, A., Kaya, H., Colakoglu, S., Bas, O., 2016. Maternal exposure to a continuous 900-MHz electromagnetic field provokes neuronal loss and pathological changes in cerebellum of 32-day-old female rat offspring. J. Chem. Neuroanat. 75, 105–110. https://doi.org/10.1016/j.jchemneu.2015.09.002.

Ogawa, K., Nabae, K., Wang, J., Wake, K., Watanabe, S.I., Kawabe, M., Fujiwara, O., Takahashi, S., Ichihara, T., Tamano, S., Shirai, T., 2009. Effects of gestational exposure to 1.95-gHz W-CDMA signals for IMT-2000 cellular phones: Lack of embryotoxicity and teratogenicity in rats. Bioelectromagnetics 30, 205–212. https://doi.org/10.1002/bem.20456.

Organization for Economic Cooperation and Development (OECD), 2007. TG 426. Developmental Neurotoxicity Study. OECD Guideline for Testing of Chemicals.

Organization for Economic Cooperation and Development (OECD), 2018. TG 414. Prenatal developmental toxicity study. OECD Guideline for Testing of Chemicals.

Orsini, N., Spiegelhalter, D., 2021. 2021; Meta-analysis of dose-response relationships, chapter 18. In: Schmid, C.H., Stijnen, T., White, I.R. (Eds.), Handbook of Meta-analysis. Publ CRC Press, Boca Raton FL USA.

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.

Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., Shamseer, L., Tetzlaff, J.M., Akl, E.A., Brennan, S.E., Chou, R., Glanville, J., Grimshaw, J.M., Hróbjartsson, A., Lalu, M.M., Li, T., Loder, E.W., Mayo-Wilson, E., McDonald, S., McGuinness, L.A., Stewart, L.A., Thomas, J., Tricco, A.C., Welch, V.A., Whiting, P., Moher, D., 2021; The PRISMA,, 2020. statement: an updated guideline for reporting systematic reviews. BMJ 372, n71. https://doi.org/10.1136/bmj.n71.

Petitdant, N., Lecomte, A., Robidel, F., Gamez, C., Blazy, K., Villégier, A.S., 2018. Alteration of adaptive behaviors of progeny after maternal mobile phone exposure. Environ. Sci. Pollut. Res. Int. 25, 10894–10903. https://doi.org/10.1007/s11356-017-1178-5.

Poulletier de Gannes, F., Haro, E., Hurtier, A., Taxile, M., Athane, A., Ait-Aissa, S., Masuda, H., Percherncier, Y., Ruffié, G., Billaudel, B., Dufour, P., Veyret, B., Lagroye, I., 2012. Effect of in utero wi-fi exposure on the pre- and postnatal development of rats. Birth Defects Res. B 95, 130–136. https://doi.org/10.1002/bdrb.20346.

Razavinasab, M., Moazzami, K., Shabani, M., 2016. Maternal mobile phone exposure alters intrinsic electrophysiological properties of CA1 pyramidal neurons in rat offspring. Toxicol. Ind. Health 32, 968–979. https://doi.org/10.1177/0748233714525497.

Rifat, F., Bhatnagara, D., Saxena, V.K., Sisodia, R., 2016. 10 GHZ microwave exposure during pregnancy causes postnatal changes in Swiss albino mice. International Journal of Pharmaceutical Sciences Review and Research 41, 75–80.

Rugh, R., McManaway, M., 1976. Anesthesia as an effective agent against the production of congenital anomalies in mouse fetuses exposed to electromagnetic radiation. J. Exp. Zool. 197, 363–367. https://doi.org/10.1002/jez.1401970309.

Rugh, R., McManaway, M., 1977. Mouse fetal sensitivity to microwave radiation. Congenit. Anom. 17, 39–45.

Rugh, R., McManaway, M., 1978. Does pre- or post-natal microwave radiation sterilize mice? Congenit. Anom. 18, 69–74.

Sambucci, M., Laudisi, F., Nasta, F., Pinto, R., Lodato, R., Altavista, P., Lovisolo, G.A., Marino, C., Pioli, C., 2010. Prenatal exposure to non-ionizing radiation: Effects of WiFi signals on pregnancy outcome, peripheral B-cell compartment and antibody production. Radiat. Res. 174, 732–740. https://doi.org/10.1667/RR2255.1.

Sangun, O., Dundar, B., Darici, H., Comlekci, S., Doguc, D.K., Celik, S., 2015. The effects of long-term exposure to a 2450 MHz electromagnetic field on growth and pubertal development in female Wistar rats. Electromagn. Biol. Med. 34, 63–71. https://doi.org/10.3109/15368378.2013.871619.

Scientific Committee on Emerging and Newly Identified Health Risks (SCENIHR), 2015. Potential Health Effects of Exposure to Electromagnetic Fields (EMF). European Commission, Luxembourg. https://ec.europa.eu/health/scientific_committees/emerging/docs/scenihr_o_041.pdf.

Schmidt, R.E., Merritt, J.H., Hardy, K.H., 1984. In utero exposure to low-level microwaves does not affect rat foetal development. Int. J. Radiat. Biol. Relat. Stud. Phys. Chem. Med. 46, 383–386. https://doi.org/10.1080/09553008414551561.

Sharma, A., Kesari, K.K., Saxena, V.K., Sisodia, R., 2017. The influence of prenatal 10 GHz microwave radiation exposure on a developing mice brain. Gen. Physiol. Biophys. 36, 41–51. https://doi.org/10.4149/gpb_2016026.

Shibkova, D.Z., Shilkova, T.V., Ovchinnikova, A.V., 2015. Early and Delayed Effects of Radio Frequency Electromagnetic Fields on the Reproductive Function and Functional Status of the Offspring of Experimental Animals. Radiats. Biol. Radioecol. 55, 514–519.

Shirai, T., Imai, N., Wang, J., Takahashi, S., Kawabe, M., Wake, K., Kawai, H., Watanabe, S., Furukawa, F., Fujiwara, O., 2014. Multigenerational effects of whole body exposure to 2.14 GHz W-CDMA cellular phone signals on brain function in rats. Bioelectromagnetics 35, 497–511. https://doi.org/10.1002/bem.21871.

Shirai, T., Wang, J., Kawabe, M., Wake, K., Watanabe, S.I., Takahashi, S., Fujiwara, O., 2017. No adverse effects detected for simultaneous whole-body exposure to multiple-frequency radiofrequency electromagnetic fields for rats in the intrauterine and pre- and post-weaning periods. J. Radiat. Res. 58, 48–58. https://doi.org/10.1093/jrr/rrw085.

Smialowicz, R.J., Ali, J.S., Berman, E., Bursian, S.J., Kinn, J.B., Liddle, C.G., Reiter, L.W., Weil, C.M., 1981. Chronic exposure of rats to 100-MHz (CW) radiofrequency radiation: Assessment of biological effects. Radiat. Res. 86, 488–505. https://doi.org/10.2307/3575465.

Smialowicz, R.J., Weil, C.M., Kinn, J.B., Elder, J.A., 1982. Exposure of rats to 425-MHz (cW) radiofrequency radiation: effects on lymphocytes. J. Microw. Power 17, 211–221. https://doi.org/10.1080/16070658.1982.11689282.

Stasinopoulou, M., Fragopoulou, A.F., Stamatakis, A., Mantziaras, G., Skouroliakou, K., Papassideri, I.S., Stylianopoulou, F., Lai, H., Kostomitsopoulos, N., Margaritis, L.H., 2016. Effects of pre- and postnatal exposure to 1880–1900MHz DECT base radiation on development in the rat. Reprod. Toxicol. 65, 248–262. https://doi.org/10.1016/j.reprotox.2016.08.008.

Takahashi, S., Imai, N., Nabae, K., Wake, K., Kawai, H., Wang, J., Watanabe, S., Kawabe, M., Fujiwara, O., Ogawa, K., Tamano, S., Shirai, T., 2010. Lack of adverse effects of whole-body exposure to a mobile telecommunication electromagnetic field on the rat fetus. Radiat. Res. 173, 362–372. https://doi.org/10.1667/rr1615.1.

Tofani, S., Agnesod, G., Ossola, P., Ferrini, S., Bussi, R., 1986. Effects of continuous low-level exposure to radiofrequency radiation on intrauterine development in rats. Health Phys. 51, 489–499. https://doi.org/10.1097/00004032-198610000-00008.

Türedi, S., Hancı, H., Çolakoğlu, S., Kaya, H., Odacı, E., 2016. Disruption of the ovarian follicle reservoir of prepubertal rats following prenatal exposure to a continuous 900-MHz electromagnetic field. Int. J. Radiat Biol. 92, 329–337. https://doi.org/10.3109/09553002.2016.1152415.

Verbeek, J., Oftedal, G., Feychting, M., van Rongen, E., Rosaria Scarfì, 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. Environ. Int. 146, 106300 https://doi.org/10.1016/j.envint.2020.106300.

Wang, L.X., Huang, H.H., Lv, G.R., 2018. cts of prenatal exposure of 850–1900MHz mobile phone on the expression of PCNA and DCX in dentate gyrus of offspring rats. Zhongguo Ying Yong Sheng Li Xue Za Zhi 34, 35–38. https://doi.org/10.12047/j.cjap.5592.2018.010.

World Health Organization (WHO), 2014. Handbook for Guideline Development, second ed. WHO Press, Geneva, Switzerland.

Wyde, M.E., Horn, T.L., Capstick, M.H., Ladbury, J.M., Koepke, G., Wilson, P.F., Kissling, G.E., Stout, M.D., Kuster, N., Melnick, R.L., Gauger, J., Bucher, J.R., McCormick, D.L., 2018. Effect of cell phone radiofrequency radiation on body temperature in rodents: Pilot studies of the National Toxicology Program's reverberation chamber exposure system. Bioelectromagnetics 39, 190–199. https://doi.org/10.1002/bem.22116.

Zhang, Y., Li, Z., Gao, Y., Zhang, C., 2015. Effects of fetal microwave radiation exposure on offspring behavior in mice. J. Radiat. Res. 56, 261–268. https://doi.org/10.1093/jrr/rru097.

Zhao, Z.G., Zhang, J.F., Yan, P., Ma, Y.M., 2005. Relationship between millimeter wave irradiation in pregnant mice and c-Fos protein expression in hippocampus and learning and memory functions in their offsprings. Zhonghua lao dong wei sheng zhi ye bing za zhi = Zhonghua laodong weisheng zhiyebing zazhi =. Chinese journal of industrial hygiene and occupational diseases 23, 424–427.

Ziskin, M.C., Morrissey, J., 2011. Thermal thresholds for teratogenicity, reproduction, and development. Int. J. Hyperth. 27, 374–387. https://doi.org/10.3109/02656736.2011.553769.

Zwetsloot, P.P., Van Der Naald, M., Sena, E.S., Howells, D.W., IntHout, J., De Groot, J.A., Chamuleau, S.A., MacLeod, M.R., Wever, K.E., 2017. Standardized mean differences cause funnel plot distortion in publication bias assessments. Elife 6. https://doi.org/10.7554/eLife.24260.