Review

Else K. Nordhagen* and Einar Flydal

WHO to build neglect of RF-EMF exposure hazards on flawed EHC reviews? Case study demonstrates how “no hazards” conclusion is drawn from data showing hazards

https://doi.org/10.1515/reveh-2024-0089
Received June 5, 2024; accepted June 25, 2024;
published online July 10, 2024

Abstract: We examined one of the first published of the several systematic reviews being part of WHO’s renewed initiative to assess the evidence of associations between man-made radiofrequency electromagnetic radiation (RF-EMF) and adverse health effects in humans. The examined review addresses experimental studies of pregnancy and birth outcomes in non-human mammals. The review claims that the analyzed data did not provide conclusions certain enough to inform decisions at a regulatory level. Our objective was to assess the quality of this systematic review and evaluate the relevance of its conclusions to pregnant women and their offspring. The quality and relevance were checked on the review’s own premises: e.g., we did not question the selection of papers, nor the chosen statistical methods. While the WHO systematic review presents itself as thorough, scientific, and relevant to human health, we identified numerous issues rendering the WHO review irrelevant and severely flawed. All flaws found skew the results in support of the review’s conclusion that there is no conclusive evidence for nonthermal effects. We show that the underlying data, when relevant studies are cited correctly, support the opposite conclusion: There are clear indications of detrimental nonthermal effects from RF-EMF exposure. The many identified flaws uncover a pattern of systematic skewedness aiming for uncertainty hidden behind complex scientific rigor. The skewed methodology and low quality of this review is highly concerning, as it threatens to undermine the trustworthiness and professionalism of the WHO in the area of human health hazards from man-made RF-EMF.

Keywords: non-ionizing radiation; electromagnetic field; environmental health; WHO

Introduction

In this paper, we present a thorough analysis of the quality, validity and conclusion of the first report of a series of reviews from a renewed World Health Organization (WHO) initiative to assess the evidence of associations between (human made) radiofrequency electromagnetic radiation (RF-EMF) and adverse health effects in the general and working population. This initiative will publish several reports in “the Environmental Health Criteria (EHC) series”. The report we analyzed [1], is the first in the row, hereinafter ‘EHC2023’, published in August 2023. A companion protocol was published in 2021 [2], below just termed “the protocol”. A monograph summing up the results of these EHC reports on individual adverse effects is planned as an update of the 1993 WHO monograph on radiofrequency fields [3].

EHC2023 is a systematic review, including meta-analyses as to 14 different endpoints related to pregnancy and birth outcomes, based on selected experimental laboratory studies on effects on pregnancy and birth from RF-EMF exposure of non-human mammals. EHC2023 is a massive work, carried out under the banner of objectivity, thoroughness and scientific rigor by the use of statistical refinement and technicalities, arguably in order to reach a sound and solid conclusion. The final conclusion of the report is (EHC2023, p. 31, bold added):

> 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.

Else K. Nordhagen and Einar Flydal contributed equally to this work.

***Corresponding author: Else K. Nordhagen**, PhD ICT, Retired Researchers, John Brandts vei 65B, 0860 Oslo, Norway,
E-mail: elsenordhagen@gmail.com. https://orcid.org/0009-0008-1257-4657
Einar Flydal, Retired Researcher & Strategy Adviser, Oslo, Norway

Our objective with this paper is, contrary to this conclusion, to document as clearly as possible, that the conclusion of the EHC2023 review is far from supported, but contradicted, by the scientific literature analyzed by its authors. What seems to be supportive, appears so due to the many irrelevant studies included, to important methodological choices not being documented or discussed, and due to significant errors and flaws. By our stepwise testing, we found that with flaws corrected, the original data and the EHC2023 methodology led to an opposite conclusion: Evidence of hazard from subthermal exposure to RF-EMF is well underpinned by its reviewed literature.

The scientific battlefield on EMF effects is heavily polarized between the researchers claiming that evidence of detrimental biological (nonthermal) health effects on humans as well as other forms of life are sufficiently well documented for immediate action, and those claiming that only hazards from acute overdue energy accumulation (overheating of tissue) are sufficiently well proven. We do position ourselves in the former camp. However, our criticism of the EHC2023 review forwarded in this paper, is first and foremost based on a wide range of flaws in the EHC2023 study which should be found inacceptable in both camps, independent of this schisma.

As far as feasible, we based our analysis on EHC2023's own premises, e.g., not questioning its selection of papers for review, neither the chosen threshold for exposures to be classified as thermal or hazardous, nor the authors' choice of statistical methodologies, e.g. basing most of their meta-analyses on the statistical concept of 'effect size', calculated in the form of Cohen's d.

As the identified flaws all skew the outcomes of the EHC2023 review in the direction of smaller effects as well as larger uncertainty about them – the quality issues seem difficult to explain as results of pure chance only. In our Discussion section we present some well-known patterns which might explain how this have come about.

Terminology

To describe the units subject to the EHC2023 review, the following terms will be used throughout this paper:

- Paper: a published paper which reports results for one or more exposure groups of animals and one or more control groups.
- Experiment: Each exposure group reported in a paper is denoted "an experiment". Experiments may, or may not, share a common control group.
- Study: an entry into a pooled effect size calculation in EHC2023 (presented pooled as forest plots). Studies usually consist of one exposure group compared to one control group, or the average from several exposure groups compared to a single, shared control group.
- Meta-study: the analysis of pooled effect size found in studies selected for an endpoint in line with the use of this term in EHC2023 and rated "Low or Some Concern" Risk of Bias.
- Thermal effects: biological effects from RF-EMF exposure caused by thermal heating of tissue.
- Thermal studies: Studies applying EMFs at intensities expected to cause effects from tissue heating.
- Nonthermal effects: Effects from RF-EMF exposure not causing thermal heating in tissue.
- Nonthermal studies/experiments: Studies/experiments of effects not caused by thermal heating.
- Thermal exposure: "High SAR" exposures with intensities assumed to cause tissue heating in the exposed animals.
- Subthermal exposure: "Low SAR" exposure of RF-EMF which is below the threshold assumed to cause tissue heating.

The EHC2023 work process and targets for our analysis

Because of the scientific complexity of EHC2023, a substantial effort is needed to scientifically evaluate the quality of the study and the relevance of the premises of its conclusions. To make the work manageable, we selected what we consider representative parts of the analysis, some with overarching consequences relevant for all meta-analysis and some focused on details in a particular endpoint, i.e. the fetal weight endpoint – the one based on the most studies. Doing so, we assumed this would give a representative picture of the EHC2023 review.

The steps of the EHC2023 procedure, which we analyzed, may be summed up as follows:

- Identify endpoints and sub-groups to analyze (done in the protocol, 14 endpoints identified).
- Identify papers on experiments on RF-EMF exposure effects on mother and fetus for the selected endpoints.
- Perform a Risk of Bias analysis of experiments, sorting them as of "High", "Some" or "Low Concern".
- Select experiments from which to collect data – one or more from each paper, forming one or more studies – for each of these studies to calculate certainty of effects from RF-EMF. (The certainty of effects is measured as "effect size" with Cohen's d with 95 % CI [Confidence Interval] for nine endpoints, as Mean Difference for two endpoints and Odds Ratio [OR] for three endpoints).

- To aggregate data from the studies to find the pooled results for each endpoint (measured as "effect size" with pooled Cohen's d with 95 % CI, pooled Means Difference or pooled Odds Ratio).
- Draw conclusions based on the pooled results from the different endpoints.

For each of the investigated 14 endpoints, a separate analysis is done for those studies rated "Low or Some Concern" and those rated "High Concern" as to Risk of Bias (RoB). In the following, we use the term "meta-study" for the result of the former as this is in line with how this term is used in EHC2023.

The authors of EHC2023 considered 88 papers to be relevant and of sufficient quality for their review. Parts of our analysis have been carried out using the data from all these 88 papers.

These 88 papers are summed up in the huge Table 3 in EHC2023, where each paper is listed with information about type of animals used and exposure conditions in the various studies, and a brief summary of each paper's findings.

For a detailed quality assessment of the performed calculations in EHC2023, we chose the procedure for doing calculations of the pooled effects size based on Cohen's d and its confidence intervals used for nine of the 14 of the endpoints. To make the work manageable, we chose the endpoint with the most studies reviewed: the "fetal weight" endpoint (EHC2023 Figure 6), based on 44 papers, assuming this would be representative for the EHC2023 meta-analyses. The assessment was done by closely reviewing 43 of these papers and how they are handled in EHC2023. The 44th paper was only available in Japanese, with an English abstract, and could therefore only be included in parts of our assessment.

Our main result: many flaws in EHC2023, all contributing to the conclusion "no conclusion can be drawn"

In this section, we present our examination of the EHC2023 in some detail. Full details are found in the Supplementary File.

More than half the studies included in EHC2023 are irrelevant for humans protected by IEEE/ICES and ICNIRP guidelines, making the results irrelevant

(For details, see Supplementary File Section A).

Since the very first guidelines (like the C95.1-1966 [4]), energy absorption thresholds for various tissues not to be exceeded have been recommended in order to avoid risk of acute hazard (i.e., thermal effects). Present IEEE/ICES and ICNIRP guidelines have similar recommendations and form the basis for radiation protection in a majority of countries. As also stated in EHC2023, (manmade) RF-EMF levels are normally well below the SAR thermal thresholds recommended by these guidelines. To our knowledge, no guidelines or praxis anywhere in the world are laxer, i.e. permitting risk for thermal effects.

IEEE/ICES and ICNIRP standards (4 and 6 W/kg, respectively), are considered to include a significant safety margin against hazards from the kind of overheating captured with dosimetry based on SAR. In EHC2023, the threshold between "High thermal SAR" and "Low nonthermal SAR" is set at 5 W/kg, a threshold which seems somewhat arbitrarily set to the average value of the IEEE/ICES and ICNIRP standards.

For the EHC2023 review to be relevant for the assessment of hazards to pregnant women and their fetuses in everyday situations, the exposure conditions in the studies reviewed should mimic – or at least to a reasonable extent be transferable to – everyday situations pregnant women will encounter when in an environment with (manmade) RF-EMFs/wireless technology. This means that only studies with "Low nonthermal SAR" are relevant.

Also, among the studies reviewed in EHC2023, there are single dose experiments. Single doses were mostly used to study effects from a brief, high SAR exposure, i.e., a brief thermal increase, a hazard also protected against in IEEE/ICES and ICNIRP guidelines and not encountered in daily and working life.

To be considered relevant, a study should therefore use subthermal exposure conditions, and over several days. We found that over half the studies selected for the EHC2023 review do not meet these criteria: They are experiments with irrelevant exposure conditions.

In the following we use the term "relevant study" for all nonthermal studies with several days of subthermal exposure. Studies/papers/experiments we found irrelevant, we just call "thermal".

In addition to being irrelevant, another problematic issue which arises from including thermal studies, is that they add heterogeneity to the meta-analyses, which means adding uncertainty to the overall result, a point also stated in EHC2023 (p. 30):

> 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 (i.e., of performing a valid and reliable meta-analysis).

As more than half of the included studies are thermal, there are therefore good reasons to question the quality and relevance of the results of EHC2023.

The many opportunities to skew the results when pooling effect sizes

When performing a meta-analysis, data are aggregated. Such a procedure of data aggregation consists of several steps where issues impacting the end result might arise. The methodology chosen by the authors of the protocol for EHC2023 has several steps which leave the results open to interpretation.

Below we first discuss two steps where issues arise: (a) how the pooled effect size lends itself to interpretation, and (b) how the choice of equation for calculating the relative weight of each study – determining how the study's result is to contribute to the pooled effect size –, may skew the result in any direction. These issues are not addressed in EHC2023 nor its protocol, something which obfuscates the results.

A pooled effect size open for interpretation

In the EHC2023, the effect size is standardized and expressed by the use of Cohen's d, a conventional, but – similar to other such statistical tools needed when aggregating data – not very intuitive. (Cohen's d is defined as "the difference between two means divided by a standard deviation for the data" [5]) Jacob Cohen, who developed this formula for calculating a pooled standardized effect size based on differences between means, suggested using the following rule of thumb for interpreting the power of the calculated effect size [6]:

- Small effect ("cannot be discerned by the naked eye") =0.2.
- Medium effect=0.5.
- Large effect ("can be seen by the naked eye")=0.8.

However, Cohen suggests caution when evaluating effect sizes, as "small" and "large" may have different importance in different contexts. Others warn that pooling effect sizes demand a certain conformity in the material pooled when suggesting that "appropriate norms are those based on distributions of effect sizes for comparable outcome measures from comparable interventions targeted on comparable samples" [7].

Thus, it follows that to conclude from effect sizes as to which of them demonstrate the strongest effects, makes sense when – and only when – comparing effect sizes from similar experiments and contexts. Such circumstances are not the case when pooling experiments investigating thermal effects with experiments investigating nonthermal effects. There are also other significant differences discussed in the Supplementary File Section D.4.2.

This observation implies that there is here a certain leeway for subjective evaluations, undermining any claims of the EHC2023 being a purely "objective assessment".

The equations used for relative weighting of the studies have the potential of skewing the results in any direction, and are neither documented nor discussed

(For details, see Supplementary File Section B).

When pooling results in meta-analyses, studies are given relative weight to mirror their importance using a weighting equation. Through the choice of equation(s), the results may be skewed in any direction. EHC2023 contains no documentation or discussion of the weighting equation(s) used in its many calculations of pooled effect sizes. However, by simulating the one used for the fetal weight endpoint, and by testing it on some of the other endpoints, we found that none of the standard weighting equations presented in the general literature have been used. Although we managed to develop a fairly close approximation (see Supplementary File), we could not find any single equation producing exactly the same results as in EHC2023. However, our search for such an equation showed us how we could skew the pooled effect size in any direction by slight modifications of the equation.

This non-transparent, non-standard weighting makes it impossible to check the quality of the weighting for possible skewedness, thereby reducing the credibility of all the EHC2023 meta-studies to a matter of trust. Without any documentation, one can only guess about the weighting, which makes the conclusions of EHC2023 questionable.

The Risk of Bias (RoB) assessment of EHC2023 is highly biased towards thermal studies

The RoB analysis is a procedure to sort studies according to their "Risk of Bias", i.e. a measurement and ranking according to quality. The assessment is made by attributing scores on a set of questions. Our analysis of the Risk of Bias questions, their interdependencies and scores attributed, demonstrates that the RoB assessment and ratings are highly

skewed toward favoring thermal, and thus irrelevant, studies, while discrediting nonthermal and relevant studies.

Almost all studies of thermal effects are also older studies, as the topic of thermal heating is by and large considered settled and protective restrictions are recommended in standards. Thus, modern RF-EMF research investigates the unsettled topic of nonthermal effects, using subthermal exposures. Hence, by favoring thermal studies, the RoB analysis is also skewed toward favoring old studies, while discrediting modern studies.

Below we explain our findings in some detail, while more is found in Supplementary File Section C:

The bias of the EHC2023 RoB analysis is easily seen by our sorting of the RoB score data, found in EHC2023 Supplementary File 5 "RoB analysis scores". Extracts from this analysis are presented below.

A summary of our findings is presented in Figure 1: we see that of the total papers reviewed, the majority is published before 2001 (50 vs. 38 papers). Of papers published 1975–2000, 22 % were rated High Concern for Risk of Bias (HC), while 52 % of the newer studies (2001–2021) are rated HC. We also see that the newer papers are almost all nonthermal, and as much as more than half of them are rated (HC), while for older studies, published before 2001, a majority of the papers are thermal, and that more nonthermal than thermal are rated HC.

The bias in ECH2023 towards favoring old and thermal studies may also be illustrated by presenting the results of the RoB analysis using the same type of graphics as is done in ECH2023 Supplementary File 5 "RoB analysis scores". When compiling these data into one table with ratings per paper, as presented in Figure 2, the bias is clearly seen just by looking at the distribution of colors: The more blue cells, means better RoB score, and the more red and orange cells, means poorer RoB score. In the table, the ranking for those questions which discriminate well between studies, are presented (questions 3.1, 3.2, 4.1, 5.1 and 6.1). The other questions do not discriminate well as they either attribute positive scores to most of the studies or negative to most of them, one even attributing positive scores to all studies (In the Supplementary File Section C, all questions are shown and discussed).

Figure 2 shows that thermal studies (yellow "Name of Paper" cells) are with just two exceptions found in the left-hand table, which lists all papers reviewed in EHC2023 published between 1975 and 2000. The thermal studies show a clear overweight of blue cells in the columns Q3.1–Q6.1, indicating that they score favorable on the specific RoB questions. There are very few red cells in the **RoB** column, indicating very few studies found to be of High Concern as to RoB.

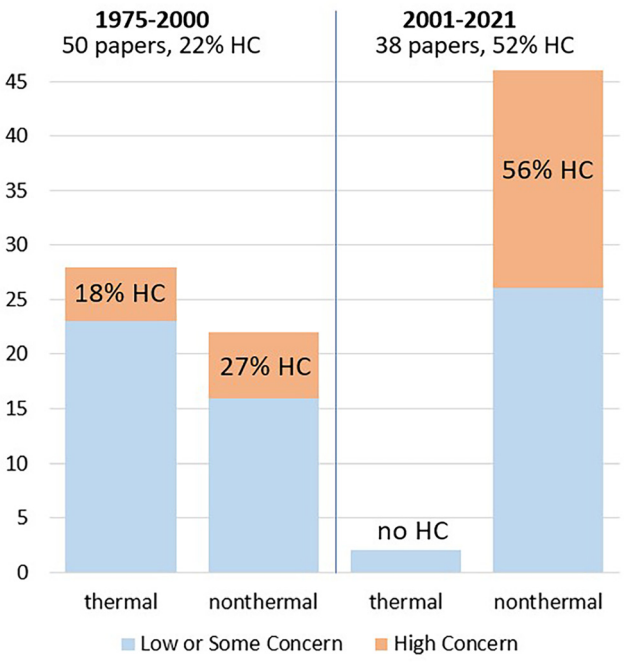

Figure 1: Distribution of older (1975–2000) and newer papers (2001–2021) and numbers of papers thermal or nonthermal for each time period. Percentages of papers rated as "High Concern" (HC) for Risk of Bias in each category are shown. The majority of papers reviewed are older studies where only 22 % are rated as HC. For newer papers 52 % are rated as HC and all of these are nonthermal studies. Among the older studies, a higher percentage of nonthermal studies are rated HC compared to thermal studies.

In the right-hand table in Figure 2, nonthermal studies published in 2001 and later are listed. Here orange and red cells dominate in the columns Q3.1–Q6.1, while. The **RoB** column shows that a majority of the studies are considered to be HC.

The downgrading of newer studies is even more pronounced when comparing papers from the last decade before EHC2023 was published, 26 papers in all, with the 26 oldest papers. Seventeen papers from the last decade are rated HC while only nine are rated SC. For the 26 oldest papers the situation is opposite: seven are HC while 19 are SC. All newer papers rated as high concern are modern, nonthermal studies using laboratory setup and measurements suited for such studies, where lack of temperature measurements and exact SAR measurements, are not signs of lack of quality.

The common feature for older studies rated HC is not lack of temperature or SAR measurements, but in all cases their unfavorable rating for Q6.1. A closer inspection shows that the low RoB rating fully correlates with few animals in the studies.

1975-2000, 50 papers

Name of paper	Pub. year	RoB	Q3.1	Q3.2	Q4.1	Q5.1	Q6.1
Chernovetz 1975	1975	HC	+ +	+ +	+ +	+	- -
Rugh 1976	1976	HC	+ +	+ +	+ +	+ +	- -
Chernovetz 1977	1977	SC	+ +	+ +	+	+ +	+
Rugh 1977	1977	SC	+ +	+	+ +	+	-
Berman 1978	1978	SC	+	+	+ +	+	+ +
Rugh 1978	1978	SC	+ +	+ +	+	+ +	-
Albert 1981	1981	HC	+	-	+ +	+	- -
Berman 1981	1981	SC	+	+	+	+ +	+ +
Nawrot 1981	1981	SC	+ +	+ +	+ +	+	+
Smialowicz 1981	1981	SC	+ +	+	+ +	+ +	+ +
Berman 1982a	1982	SC	+	+	+	+ +	+
Berman 1982b	1982	SC	+	+ +	+	+ +	+ +
Jensh 1982a	1982	SC	+	+ +	+ +	+ +	+
Jensh 1982b	1982	HC	+	+ +	+	+ +	- -
Kaplan 1982	1982	HC	+ +	+ +	+	+	—
Lary 1982	1982	SC	+	+ +	+ +	+	+ +
Smialowicz 1982	1982	SC	+ +	-	-	+ +	-
Chazan 1983	1983	SC	+ +	+	+ +	+	+
Galvin 1983	1983	HC	+ +	+ +	+	+ +	- -
Inouye 1983	1983	SC	+ +	+	+ +	+	-
Jensh 1983a	1983	HC	+	+ +	+ +	+ +	- -
Jensh 1983b	1983	SC	+	+ +	+	+ +	+
Lary 1983a	1983	SC	+ +	+ +	+ +	+	+ +
Lary 1983b	1983	SC	+	+ +	+ +	+ +	+
Berman 1984a	1984	SC	+	+ +	+ +	+ +	+ +
Berman 1984b	1984	SC	+	+	+	+ +	+ +
Jensh 1984a	1984	SC	+	+ +	+ +	+ +	+
Jensh 1984b	1984	SC	+	+ +	+	+ +	+
Merritt 1984	1984	SC	+ +	+	+ +	+ +	-
Schmidt 1984	1984	SC	+ +	+	+ +	+ +	+
Nawrot 1985	1985	SC	+ +	+ +	+ +	+	+
Galvin 1986	1986	SC	+ +	+	+ +	+	+
Lary 1986	1986	SC	+	+ +	+ +	+	+
Marcickiewicz 1986	1986	SC	+	+	-	+	-
Tofani 1986	1986	SC	+	+ +	+	+ +	+ +
Brown-Woodman 1988	1988	HC	-	+ +	+ +	+ +	- -
Brown-Woodman 1988	1988	SC	+	+ +	+ +	+ +	-
Chiang 1988 (CS, 3xHC)	1988	SC	-	+	-	+	-
O'Connor 1988	1988	HC	-	+	+	+	- -
Nelson 1991	1991	SC	+	+ +	+ +	+ +	+ +
Berman 1992	1992	HC	+ +	- -	+	+ +	+ +
Nelson 1994	1994	SC	+	+ +	-	-	-
Kubinyi 1996	1996	SC	+ +	-	+	+ +	+ +
Inaloz 1997	1997	HC	- -	+ +	+ +	+	-
Nelson 1997a	1997	SC	+ +	+ +	-	-	-
Nelson 1997b	1997	SC	+	+ +	-	-	+
Nelson 1999	1999	SC	+	+	+	+	-
Bornhausen 2000	2000	SC	+	+	+	+	+ +
Cobb 2000 (3xSC, 3xHC)	2000	SC	+	+	+	+ +	-
Dasdag 2000	2000	SC	-	+ +	+ +	+	-

2001-2021, 38 papers

Name of paper	Pub. year	RoB	Q3.1	Q3.2	Q4.1	Q5.1	Q6.1
Nelson 2001	2001	SC	+	+ +	+	+	-
Anderson 2004	2004	LC	+ +	+	+ +	+ +	+ +
Zhao 2005	2005	HC	+	+	+ +	+ +	-
Ferreira 2006	2006	HC	+	+	+ +	+ +	- -
Lee 2009	2009	SC	+ +	+ +	+ +	+ +	+
Ogawa 2009	2009	SC	+ +	+	+ +	+	+ +
Guler 2010	2010	SC	-	-	-	+	-
Sambucci 2010	2010	LC	+ +	+	+ +	+ +	+
Takahashi 2010	2010	SC	+ +	+	+ +	+ +	+
Ait-Aissa 2012	2 012	SC	+ +	-	+ +	+	+
Aldad 2012	2012	HC	- -	-	+ +	+	+ +
Poulletier de Gannes 2012	2012	LC	+ +	+	+	+ +	+
Bas 2013	2013	HC	- -	-	-	+ +	- -
Haghani 2013 2xHC	2013	HC	- -	+ +	-	+ + (1, "+" 2)	-
Ikinci 2013	2013	HC	+	-	+	+	-
Odaci 2013	2013	HC	-	-	-	-	- -
Shirai 2014	2014	SC	+ +	+	+ +	+ +	-
Sangun 2015	2015	HC	+	+	-	+ +	- -
Shibkova 2015	2015	HC	- -	- -	+	+	+
Zhang 2015	2015	SC	-	-	-	+ +	-
Alchalabi 2016	2016	HC	- -	- -	-	+	+
Erdem-Koc 2016	2016	SC	-	-	+ +	+	-
Odaci 2016	2016	SC	-	-	-	+	-
Razavinasab 2016	2016	HC	-	-	-	+	- -
Rifat 2016	2016	HC	+	-	-	-	-
Stasinopoulou 2016	2016	SC	+	+	+	-	-
Türedi 2016	2016	HC	-	-	+	+ +	- -
Alchalabi 2017	2017	HC	- -	- -	+	+	+ +
Sharma 2017	2017	SC	+	-	-	+	-
Shirai 2017	2017	SC	+ +	+	-	-	+
DastAmooz 2018	2018	HC	- -	-	-	+ +	-
Petitdant 2018	2018	HC	+ +	-	-	-	-
Wang 2018	2018	HC	- -	-	+ +	+	-
Wyde 2018	2018	SC	+ +	+	+	-	+
Calis 2019	2019	SC	-	-	-	+	+
Azimzadeh 2020	2020	HC	- -	-	-	+	+
Li 2020	2020	HC	- -	-	+	+ +	- -
Keles 2021	2021	HC	+	+	-	+	-

Figure 2: The EHC2023 RoB ranking for questions which discriminate well between studies, with same color coding as in EHC2023. Left table: rankings for the papers reviewed published year 2000 and earlier. Right table: papers published from 2001 onwards. Yellow background in "Name of Paper" cells: thermal papers. Such papers dominate among older papers (left) while almost completely absent among more recent papers (right). Risk of Bias (RoB) columns: "High Concern" (HC) overall rating of paper shown in red. Most more recent studies are rated HC, while hardly any of the older, mostly thermal, studies. RoB questions Q3.1–Q6.1: Shades of blue dominate in left table, showing majority of the elder papers get positive RoB ratings (i.e., low risk). Orange and red dominate to the right, showing that the majority of papers from 2001 onward, in particular new ones published the last decade before WHC2023 was published, get negative ratings for these questions. This trend is even more pronounced when comparing the 26 newest with the 26 oldest papers (thick borders): 17 papers published from 2013 and onward are rated HC while only 9 are rated SC. For the 26 oldest papers the situation is opposite: 7 are HC while 19 are SC.

The overall conclusion from this analysis is that the RoB analysis is biased, in particular as it downgrades modern studies of nonthermal effects due to RoB questions unfit for quality assessment of such studies.

Although pooled effect sizes are calculated separately in EHC2023 both for papers with "Low or Some Concern" (LSC) and papers with "High Concern" (HC), the conclusions are mainly drawn on the pooled effect size calculated from the LSC studies, i.e. including the thermal papers. The very clear bias of the RoB analysis causes the meta-analyses for all endpoints and thus the overall conclusion of EHC2023, to be based largely on thermal studies irrelevant to humans protected by present thermal based RF-EMF regulations.

The pooled effect size of the HC studies, where nonthermal studies dominate, is stronger than the pooled effect sizes of the LSC studies. Accepting the nonthermal studies would therefore increase the effect size. Their higher effect sizes are in EHC2023 "explained away" in a curious case of circular reasoning:

The authors of EHC2023 assume that the higher effect sizes found in HC papers are caused by the authors exaggerating their findings. This assumption is set forward

without any foundation, and could just as well be that the newer, nonthermal studies used modern methods more apt at detecting nonthermal effects. Thus, the assumption seems a reflection of the SAR-based tradition of which the entire EHC undertaking is rooted, questioning the existence of nonthermal effects.

In the next run, this biased reasoning is extended to downplaying the many detrimental effects in fact found in the "Some Concern" and "Low Concern" nonthermal papers in the EHC2023 (presented here in later section): This downplaying is sought underpinned by inferring that since authors of non-thermal "High Concern" papers (purportedly) exaggerate their results, this could also to be the case for nonthermal "Some Concern" papers. Further on, this argument is extended, by inverting the logic, into downplaying all nonthermal effects found by speculating that the RoB analysis resulted in very few "Low Concern" nonthermal papers because the authors of such papers generally exaggerate their findings:

> the evidence of an (nonthermal) RF-EMF effect on pregnancy and birth outcomes could be overstated to a certain degree (EHC2023, p. 21).

The biased RoB analysis of the EHC2023 has the overall effect of increasing the number of thermal studies used in the meta-analysis – thereby reducing relevance, increasing heterogeneity and adding uncertainty, and is then further used in a biased and flawed line of logic to claim that the resulting small pooled effect sizes then found are uncertain.

Our in-depth analysis of the "fetal weight" meta-analysis reveals misciting and miscalculation behind the pooled effect size

As mentioned above, we chose, for practical reasons, the "fetal weight" endpoint for an in-depth analysis. This endpoint meta-analysis is riddled with errors and flaws having significant impacts on the calculated effect size, and is strongly miscommunicating significant risks reported. For details see Supplementary File Section D.

Due to the biased RoB analysis, more than half (54 %) of the experiments in the studies selected in EHC2023 for this endpoint, are thermal and with their focus on investigating effects of heating on the pregnant mothers and fetuses. The results from the EHC2023 fetal weight meta-analysis should already at this point be considered irrelevant as to assessing health risks of women and their fetuses in daily life and work conditions.

To make our in-depth analysis relevant, we limited it to the nonthermal studies listed in EHC2023 Figure 6, which include studies with LSC as well as HC RoB rating, yielding 33 out of 62 studies to be analyzed in detail.

The numerous flaws all downplay effects found in the studies, thus creating the false impression of few and minor effects, lowering the quality of the meta-analysis to – in our opinion – an unacceptable level.

We identified the following flaws:

(1) Errors in the reproduction of data from the studies reviewed: Out of the 33 relevant studies, we found five studies cited erroneously. E.g., number of dams is doubled; weight of exposed fetuses is cited as 0.94, while 0.89 in the study; the effect size is made 10 times larger by mis-citing the average standard deviation; control and exposed groups are mixed up so that the study is cited as showing fetal weight increased from exposure, while the paper reports significantly reduced fetal weight (Supplementary File Section D.1).
(2) Undocumented use of variations in equations affecting confidence interval (CI) values: The calculations used in EHC2023 affecting the confidence (CI) intervals, have been done in a deviant way for the four studies with the fewest animals, attributing these studies a higher impact relative to other studies, without any reasons given (Supplementary File Section D.2).
(3) "Cherry picking" of experiments from which data are gathered: Handling data from studies with experiments sharing a common control group, relevant for 12 studies, should be done by choosing the methods best fit to reflect the most relevant findings reported in the paper. However, in the fetal weight meta-analysis, the choices are mostly undocumented, and all without scientifically valid arguments for why the authors picked one experiment over the other. We find that all the selections made contribute to reducing the effect size compared to the obviously more relevant alternative choices. E.g., EHC20203 chooses only to report weight for male fetuses, although female fetuses have more reduced weights for exposed groups, i.e. higher effect size. Such choices leave a clear impression of "cherry picking": Those experiments instrumental to favor the conclusion "no pooled effect found" are selected (Supplementary File Section D.4).

We found, in total nine studies (27 %) cited erroneously or with unjustified divergences, and "cherry picking" experiments from 12 studies (36 %). In some cases, several flaws applied to the same study. In total we found flawed handling of 19 (58 %) of the 33 nonthermal studies analyzed in EHC2023 for the fetal weight endpoint.

All of the flaws described skew the results in the direction of a lower pooled effect size and/or higher uncertainty. The flaws make the meta-analysis useless, unless for just that purpose.

A simple example was found when checking the pooled effect size for the HC (High Concern) studies, where EHC2023 reports a value of 0.52. When values are correctly reproduced and the same equation used for all studies (flaws type 1 and 2 above, Supplementary File Section D.3), the result is 0.83 and with a narrower CI (Confidence Interval). When applying the standard interpretation of Cohen's d, 0.83 is both a significant more certain finding and a stronger detrimental effect.

Misreporting of significant detrimental effects found in reviewed papers due to choice of endpoints not used in modern, nonthermal studies

(For details, see Supplementary File Section E).

EHC2023 Table 3 offers an overview of findings from the experiments. To be of acceptable quality, this overview should give a correct impression of the findings actually reported in the papers reviewed. To check the quality of the overview, we examined how findings in the papers are reported. For capacity reasons, we limited our analysis to the 28 relevant, nonthermal papers selected for the EHC2023 fetal weight meta-analysis. We found that significant detrimental effects were not reported, or misreported in various ways:

Of the 28 papers, eight report no effects found. Of the remaining 20 papers, significant results are reported in 13 papers (65 %) but omitted in EHC2023 Table 3 (A very few of these omitted effects are superficially mentioned elsewhere in EHC2023).

Only for the remaining seven of the 28 nonthermal papers are all effects reported in EHC2023 Table 3, although they are still downplayed in most cases.

The severe findings unreported are mostly physiological changes, such as changes in biochemical parameters common to all mammalian offspring (enzymes, free radicals, blood values, gene damage) and effects on cell structures (brain, nerves). Hence, all these effects should be expected to appear in all mammal offspring – rats, mice and humans alike – and can be detrimental. Therefore, they are highly relevant to the assessment of health risks. Allegedly, these endpoints are not reported in Table 3, as the authors of EHC2023 did not find them important, since only "a very brief description of the main results in scope for the systematic review is reported, based on the authors' interpretation and discussion" (EHC2023, p. 8). This statement seems to imply that "the scope of the EHC review" and of the EHC2023 "defines away" significant and relevant findings. So, what is this scope?

The scope, i.e. the endpoints to study, are defined in the protocol, listing pregnancy outcomes and endpoints prescribed in OECD TG 414 and 426, none of which address physiological changes. Why so? Again, the answer seems linked to bias favoring old studies and old thinking:

Using biochemical parameters as endpoints is part of modern approaches to assess health risks from subthermal exposure. Since such effects are relevant to all mammals, they offer better assessments also as to human health risks when experiments are carried out on animals. By disregarding findings at such endpoints and downgrading such modern approaches in their RoB assessment, the authors of EHC2023 favor old, irrelevant studies. Also, by their endpoints chosen, they effectively eliminate the huge majority of modern research observing physiological changes clearly caused from subthermal exposures [8].

Hence, EHC2023 Table 3 conveys a false impression of few and uncertain findings of detrimental effects, in stark contrast to the multitude of significant detrimental effects reported in the papers reviewed.

Relevant correlated factors not taken into account

(For details, see Supplementary File Section F).

In any causal or correlation analysis, not taking significant, widely known and accepted correlating factors into account means making the result uncertain and highly questionable. Our checks of the EHC2023 meta-analyses of both "fetal weight" and "brain weight" endpoints revealed such errors, making the resulting analyses in EHC2023 on the effects of RF-EMF too simple and shallow:

EHC2023 checks for fetal weight reduction as an effect from RF-EMF exposure. However, the well-known inverse correlation between fetal weight and litter size has not been taken into account by the EHC2023 authors, although the topic is discussed in several of the papers reviewed: The more fetuses in one litter, the lower the individual weight – independently of the RF-EMF exposure.

So, we did a simple correlation check for these parameters in the 28 studies reporting both on litter size and fetal weight. We found that the overall result does indeed support the most common effect from RF-EMF exposure to be reduced fetal weight. We thereby strengthened the certainty of that finding.

We also checked for the "brain weight" endpoint: The EHC2023 meta-analysis produces the result "no effect found".

However, for natural reasons, brain weight correlates with fetal weight which again correlates with litter size. None of these correlations are addressed.

In addition, EHC2023 presents reduced brain weight as being a good indicator of physiological changes in the brain (p. 24). This is a severe error: The studies reviewed in EHC2023 do find significant physiological changes resulting from RF-EMF exposure, however while brain weight is not reduced, or – to the contrary – even increased. Thus, basing a final assessment on brain weight, as EHC2023 does, equals to skewing the result towards "no effect" as to significant physiological changes.

It Is alarming that the two single endpoints in EHC2023 that we checked, are so "saturated" with serious flaws, rendering the meta-analyses useless, although both endpoints should have been rather simple to evaluate. The lack of reasonable attention to these endpoints' correlations – with each other as well as with other possibly relevant factors – underpins that a more solid analysis should have been carried out. This flaw alone makes the meta-analyses misguiding and useless, unless, here again, for arriving at the conclusion that "no conclusions can be drawn".

Sub-group analyses of non-thermal effects not made, although prescribed in protocol

(For details, see Supplementary File Section G).

The protocol for EHC2023 specifies that sub-group analyses should be made for subthermal RF-EMF exposure – for parameters such as frequency, duration and modulation (the protocol, p. 9). EHC2023 deviates from its protocol by only doing sub-group analysis for animal species, SAR-values (SAR<0.1, 0.1≤SAR<5, SAR≥5 W/kg) and animal core temperature increase. The reason for leaving out analyses of effects from such important properties of RF-EMF subthermal exposure is stated in (EHC2023 p. 5): "because they (i.e. animal species, SAR-values and animal core temperature increase) were considered the most likely to affect a possible association between exposure and outcomes and to keep the work manageable."

EHC2023 refers (ibid) to this important omission as "a slight deviation from the protocol", a comment that reflects the authors' bias due to "thermal thinking". In fact, many of the reviewed papers find significant variation in effects from subthermal exposures with varying frequency, duration and/or modulation.

A further reason stated in EHC2023 to leave out such important subgroup analyses prescribed by the protocol is lack of studies, as the protocol, purportedly, requires at least three studies in each subgroup (ibid, p. 9). However, nowhere in the protocol is lack of studies mentioned as a reason not to do such subgroup analyses, and, also, there would be no lack of studies, where it not for EHC2023 selecting outdated endpoints and demanding irrelevant dosimetry [8].

Leaving out sub-group analyses for frequency, duration and modulation means missed opportunities to spot effects on health from subthermal exposure and their possible causes, as such analyses would do so better than sub-groups for averaged SAR values from heterogeneous experiments.

When addressing regulatory bodies, EHC2023 downplays hazards it identified

(For details, see Supplementary File Section H).

For the practically minded, the most important section of EHC2023 is 4.4. Implications for policy and research (p. 31), where advice is given to regulators of RF-EMF exposure.

From just this section, the readers are led to believe that no change in regulation is needed, since, no conclusions can be drawn from the reviewed papers. Only endpoints with no or very small and uncertain effect sizes found are mentioned, and even those findings are downplayed. The moderate and large effect sizes found in EHC2023 are not mentioned:

For example, in the category named "Delayed effects on the offspring health". It is clearly stated in EHC2023 (p. 39) that most of these five endpoints show moderate to large detrimental effect size. Due to their relevance with respect to human offspring, these findings are of utmost relevance as to the aim of the systematic review. Still, when advising regulators, the authors do not mention these findings – nor do they mention the need for precautionary measures. This amounts to what seem an act of irresponsibility, even when evaluated on the review's own premises.

EHC2023 claims the effects found stem from the thermal studies. But when corrected, the strongest effects are in the nonthermal

(For details, see Supplementary File Section I).

To demonstrate the consequences of the many flaws revealed by our examination of the EHC2023 "fetal weight" meta-analysis, we did a stepwise recalculation of the pooled effect size and confidence interval (CI) using correct and relevant data from the relevant studies.

All along, we used the same statistical methods as used by the WHO20203 authors.

First, we corrected for bias in the EHC2023 RoB analysis, by only including relevant studies:

(1) The pooled effect size got significantly weaker when corrected for bias in the Risk of Bias analysis by including all nonthermal studies (i.e., with SAR below 5 W/kg), i.e., also those classified as "High Concern", and by excluding all thermal studies: The pooled effect size went down from 0.35 to 0.24 (C in Figure 3).
Secondly, we used the same studies as point one above, while correcting for errors and bias in experiment selection:

(2) The pooled effect size increased significantly – from 0.24 up to 0.78 – when recalculated after correcting errors (i.e. using data as reported in the studies) and selecting the experiments most relevant to humans in studies sharing a control group (corrections in 19 of 33 studies). (See [D] in Figure 3) In addition to an increased effect size, more certainty that subthermal exposure has detrimental effects was added as the CI got narrower.
Thirdly, we extracted the experiments with SAR below 2 W/kg only:

(3) The pooled effect size increased significantly – from 0.78 up to 1.01 – when calculated only from experiments using SAR below 2 W/kg (E in Figure 3).

This result shows that of all the nonthermal studies in the EHC2023 fetal weight analysis, those with the lowest SAR show the highest reductions of fetal weight and therefore contribute the most to the increase in pooled effect size for the "fetal weight" endpoint. It is also significantly higher than the pooled effect size presented in EHC2023 for the thermal, LSC studies (1.01 compared to 0.51).

It seems unavoidable to conclude that the EHC2023 claim cited at the beginning of this subsection is not just misleading, but truly false, due to the significant errors and flaws in how the experiments and their data have been handled. In addition, the claim is also bewildering and obfuscating by suppressing that effects from nonthermal studies are significant and strong when data are correctly reported and just criteria and calculation methods are used.

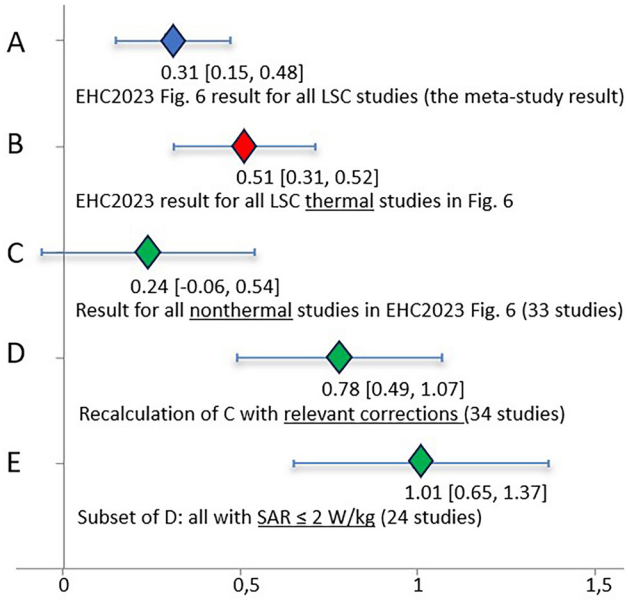

Figure 3: Forest plot for "fetal weight" endpoint of pooled effect sizes (Cohen's d) and confidence intervals [CI], demonstrating stepwise the effects of correcting flaws found in EHC2023, with A and B included for comparison. (A, blue diamond) The result, i.e., the pooled effect size, as calculated in EHC2023 Figure 6 for studies of "Low and Some Concern (LSC)" RoB. This is the result for the EHC2023 meta-analysis for the endpoint "fetal weight". (B, red) The result for LSC thermal studies only, as presented in EHC2023 Figure 6. (C, D, E; green) Results from our three calculations and CI for relevant, nonthermal studies (SAR<5 W/kg), independently of RoB scores; (C) result for all the subthermal studies, with values as in EHC2023 Figure 6, i.e. no corrections done. (D) Result of recalculations for the same studies as in (C), where flaws are corrected, rising pooled effect size from 0.24 to 0.78. (E) Recalculated results for the subset of studies in D, where SAR≤2 W/kg, raising pooled effect size to 1.01, an effect size which, according to Cohen, is clear evidence of an effect.

Discussion: how may all the flaws in EHC2023 be explained?

Were it not for the many flaws in data rendering, inconsequential and obscure cherry-picking and weighting, and the misreporting and omissions of damage found in the reviewed studies, EHC2023 could be considered an interesting review founded on well-established biophysics and toxicology. It could also be considered particularly trustworthy, as it hails strict demands for scientific proof, basing itself on objective criteria – i.e., observable endpoints in terms of substantiated damage, well understood causation, and on a straightforward dosimetric relationship well established: energy intensity thresholds resulting in thermal damage when surpassed. It would, without the flaws, also contribute to a better overview and acceptance of subthermal effects.

However, EHC2023's many flaws strip the study of legitimacy and the results of any credibility or trustworthiness, even when evaluated on the premises of the so-called "thermal only tradition".

On the top of this, the authors of EHC2023 falsely claim that all effects found be caused by results of experiments

with thermal exposure conditions and therefore, falsely inferring that realistic conditions are harmless.

A substantial part of the flaws may well be explained as resulting from all the authors of the review belonging to the "the thermal only tradition":

The WHO EHC undertaking is organized by a group of 21 experts selected by WHO's office "The International EMF Project" (TIEP). All but three of these experts are either affiliated with the guidelines issuing foundation ICNIRP [9] or connected through co-authorships linked to a tiny self-referencing network of authors mainly affiliated with ICNIRP and IEEE/ICES or both [10]. Both ICNIRP and IEEE/ICES defend exposure limits based on tissue overheating as sufficient for EMF protection, the thermal only tradition. ICNIRP is closely connected to TIEP. IEEE/ICES issues guidelines for RF-EMF used in USA, in co-ordination with ICNIRP.

The 21 experts selected the authors for the protocol and EHC2023. All these experts are proponents of the thermal only view.

Hence, the bias found in the Risk of Bias (RoB) analysis and the choice of endpoints could be considered heritage of the outdated thermal only tradition. Also, although under considerable doubt, we could imagine that even the "cherry picking" of experiments and the misreporting of hazardous effects could be interpreted as the unconscious results of strong confirmation bias. And, in its own right, the EHC2023 authors' reasoning that nonthermal findings – more than thermal ones – are being exaggerated, could be regarded as a textbook example of confirmation bias and circular reasoning deriving from the mindset of a thermal only tradition.

On the other hand, the outright false and systematic skewed selection of data from the reviewed studies, and the omissions of clearly relevant findings of hazard relevant to humans, cannot be explained as a result of confirmation bias or a blindfolding thermal only tradition. And, as the EHC2023 is a thoroughly done analysis, not a hastily made report, they are not chance errors, as they all skew the result towards a thermal only conclusion and are essential for getting such a result. Hence, the flaws seem the results of a highly intentional act.

Neither is the omission of documenting the weighting equation(s) used in the meta-analysis explainable by tradition or bias: The omission renders the meta-analysis uninterpretable as to possible health risks, as it is impossible to validate if the equation(s) used are biased or balanced in respect to EHC2023's proclaimed goal of assessing health risks in real life situations.

Conclusions

The rigorous protocol and extensive analyses presented in EHC2023 and its protocol, convey an impression of serious science, credibility, and reliability. However, we have shown that this is not the case.

We found EHC2023 to be a massive work with a rigorous and complex protocol and extensive and complex statistical analyses. A consequence of the complexity is that it can be assumed that no average reader – not even professionals – will check the results of the review, if not for other reasons, because of the major effort needed. Thereby, scientific exchange, debate and control is impeded and reduced to a matter of trust.

We had the opportunity to spend time on an in-depth analysis of representative parts of EHC2023 to assess its quality based to the extent possible on the review's own premises – that is, independent of our opinion about the professional premises chosen.

We cannot prove that the flaws and omissions are deliberately added to reach wanted conclusions, as we have next to no information about the authors, neither of the process behind the authoring EHC2023 or its protocol. Anyhow, and whatever the cause, the EHC2023 review is clearly of such a low quality, also when evaluated within the thermal only tradition, that its conclusions are without scientific value.

Our findings show that the conclusion of EHC2023 is not well-founded, and therefore the final conclusions of EHC2023 that no conclusion can be drawn that are (EHC2023, p. 31) "certain enough to inform decisions at a regulatory level" cannot to be trusted. The errors, flaws and omissions are grave enough to render EHC2023 unscientific and unethical, and it should therefore be retracted.

As it now stands, the conclusion of ECH2023 stands out as what appears to be a manufactured argument for current regulations being adequate to protect the health of human mothers and their offspring. Manipulating and skewing research results in order to manufacture a wanted conclusion is a well-known strategy to avoid stricter regulations [11–15]. Further investigations and better sources would be needed to prove such an assault on humanity to be the case as to EHC2023.

EHC2023 is just one of several studies commissioned by the same organization (WHO EHC no. 137) and states clearly that consistency has been assured in the protocols for these studies. Our analysis of EHC2023 may in this view be seen as a case study of the results of the entire WHO EHC undertaking: Since many of our concerns are related to core

elements of the protocol, there are good reasons also to question the quality of all present and forthcoming results being part of the WHO EHC undertaking unless a thorough revision of its course is made.

Research ethics: Not applicable.
Informed consent: Not applicable.
Author contributions: The authors have accepted responsibility for the entire content of this manuscript and approved its submission.
Competing interests: The authors state no conflict of interest.
Research funding: None declared.
Data availability: Not applicable.

References

1. Cordelli E, Ardoino L, Benassi B, Consales C, Eleuteri P, Marino C, et al. Effects of radiofrequency electromagnetic field (RF-EMF) exposure on pregnancy and birth outcomes: a systematic review of experimental studies on non-human mammals. Environ Int 2023;180: 108178.
2. Pacchierotti F, Ardoino L, Benassi B, Consales C, Cordelli E, Eleuteri P, et al. 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 2021;157:106806.
3. World Health Organization. Electromagnetic fields (300 Hz to 300 GHz), Environmental health criteria (EHC) series #137. Geneve: WHO; 1993. Available from: http://www.inchem.org/documents/ehc/ehc/ehc137.htm.
4. ANSI C95.1-1966. American national standard safety level of electromagnetic radiation with respect to personnel. New York, NY: Institute of Electrical and Electronics Engineers; 1966.
5. Wikipedia. Effect size. 2023. Available from: https://en.wikipedia.org/wiki/Effect_size#Cohen's_d.
6. Cohen J. Statistical power analysis for the behavioral sciences. London, UK: Routledge; 1988. ISBN 978-1-134-74270-7.
7. Lipsey MW, Puzio K, Yun C, Herbert MA, Steinka-Fry K, Cole MW, et al. Translating the statistical representation of the effects of education interventions into more readily interpretable forms. Washington, DC: U.S. Dept of Education, National Center for Special Education Research, Institute of Education Sciences. NCSER 2013–3000; 2012.
8. Lai H. Research summaries (may 2022). 2022. BioInitiative Report. Available from: https://bioinitiative.org.
9. Moskowitz J. WHO radiofrequency EMF health risk assessment monograph (EHC series) – WHO announces members of the task group on radiofrequency fields and health risks. 2023. Available from: https://www.saferemr.com/2021/09/who-radiofrequency-emf-health-risk.html.
10. Nordhagen EK, Flydal E. Self-referencing authorships behind the ICNIRP 2020 radiation protection guidelines. Rev Environ Health 2022; 38:531–46.
11. Legg T, Hatchard J, Gilmore AB. The science for profit model – how and why corporations influence science and the use of science in policy and practice. PLoS One 2021;16:e0253272.
12. Rampton S, Stauber J. Trust us, we're experts!. New York, NY: Putnam; 2001.
13. Conway EM, Oreskes N. Merchants of doubt: how a handful of scientists obscured the truth on issues from tobacco smoke to global warming. London, UK: Bloomsbury Publishing; 2012.
14. Markowitz G, Rosner D. Deceit and denial – the deadly politics of industrial pollution. California: Univ. of California Press; 2002.
15. Michaels D. Doubt is their product: how industry's assault on science threatens your health. Oxford, UK: Oxford; 2008.

Supplementary Material: This article contains supplementary material (https://doi.org/10.1515/reveh-2024-0089).