Environment International 161 (2022) 107106

Contents lists available at ScienceDirect

Environment International

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

Effects of radiofrequency electromagnetic fields (RF EMF) on cancer in laboratory animal studies

Meike Mevissen [a,*], Jerrold M. Ward [b], Annette Kopp-Schneider [c], James P. McNamee [d], Andrew W. Wood [e], Tania M. Rivero [f], Kristina Thayer [g], Kurt Straif [h,i]

[a] Veterinary Pharmacology & Toxicology, Department of Clinical Research and Veterinary Public Health (DCR-VPH), University of Bern, Bern, Switzerland
[b] Global VetPathology, Montgomery Village, Maryland, USA
[c] Division of Biostatistics, German Cancer Center, Heidelberg, Germany
[d] Non-Ionizing Radiation Health Sciences Division, Consumer and Clinical Radiation Protection Bureau, Health Canada, Ottawa, Canada
[e] Department of Health Sciences and Statistics, Swinburne University of Technology, Hawthorn, Australia
[f] Medical Library, University Library, University of Bern, Bern, Switzerland
[g] Center for Public Health and Environmental Assessment, Chemical & Pollutant Assessment Division, US EPA, NC, USA
[h] ISGlobal, Barcelona, Spain
[i] Boston College, MA, USA

ARTICLE INFO

Handling Editor: Adrian Covaci

Keywords:
Radiofrequency electromagnetic fields
Carcinogenicity
Bioassay
Animal studies
Carcinogenesis, toxicity, systematic review
Radiofreuency exposure

ABSTRACT

Background: The carcinogenicity of radiofrequency electromagnetic fields (RF EMF) has been evaluated by the International Agency for Research on Cancer (IARC) in 2011. Based on limited evidence of carcinogenicity in humans and in animals, RF EMF were classified as possibly carcinogenic to humans (Group 2B). In 2018, based on a survey amongst RF experts, WHO prioritized six major topics of potential RF EMF related human health effects for systematic reviews. In the current manuscript, we present the protocol for the systematic review of experimental laboratory animal studies (cancer bioassays) on exposure to RF fields on the outcome of cancer in laboratory animals.

Objective: In the framework of WHO's Radiation Program, the aim of this work is to systematically evaluate effects of RF EMF exposure on cancer in laboratory animals.

Study eligibility and criteria: WHO's Handbook (2014) for guideline development will be followed with appropriate adaptation. The selection of eligible studies will be based on Population, Exposures, Comparators, and Outcomes (PECO) criteria. We will include peer-reviewed articles and publicly available reports from government agencies reporting original data about animal cancer bioassays on exposure to RF EMF. The studies are identified by searching the following databases: MEDLINE (PubMed), Science Citation Index Expanded and Emerging Sources Citation Indes (Web of Science), Scopus, and the EMF Portal. No language or year-of-publication restrictions are applied. The methods and results of eligible studies will be presented in accordance with the PRISMA 2020 guidelines.

Study appraisal method: Study evaluation of individual studies will be assessed using a risk of bias (RoB) tool developed by the Office of Health Assessment and Translation (OHAT) with appropriate considerations including sensitivity for evaluating RF EMF exposure in animal cancer bioassays. The final evaluation on the certainty of the evidence on a carcinogenic risk of RF EMF exposure in experimental animals will be performed using the OHAT *Grading of Recommendations Assessment, Development and Evaluation* (GRADE) approach with appropriate considerations.

The protocol has been registered in an open-source repository (PROSPERO).

Funding: The study is partly financially supported by the World Health Organization. No additional funding was provided outside author salaries through their places of employment.

* Corresponding author.
E-mail address: meike.mevissen@vetsuisse.unibe.ch (M. Mevissen).

https://doi.org/10.1016/j.envint.2022.107106
Received 6 July 2021; Received in revised form 3 December 2021; Accepted 18 January 2022
Available online 29 January 2022

1. Introduction

1.1. Background

The technological applications of radiofrequency electromagnetic fields (RF EMF; frequencies 100 kHz to 300 GHz) have been steadily increasing since the 1950s. RF EMF are used in medical diagnostics and therapy (e.g., magnetic resonance imaging (MRI), diathermy, radiofrequency ablation), industry (e.g., heating and welding), domestic appliances (e.g. baby monitor, Wi-Fi), security and navigation (e.g., radar), and especially telecommunications (radio and television, mobile phones and wireless networks). Both environmental and occupational exposure to RF EMF have increased and are further increasing; the roll-out of 5G will add new frequencies and the prevalence of exposure will further increase. Concern has been raised regarding public health consequences from RF EMF, and it is therefore crucial to perform a health risk assessment to support decision-makers and inform the general public.

The World Health Organization (WHO) has an ongoing project to assess potential adverse health effects of exposure to RF EMF in the general and working population. To prioritize potential adverse health outcomes from exposure to these fields, WHO conducted a broad international survey amongst RF experts in 2018 (Verbeek et al., 2021). Six major topics were identified for which WHO has now commissioned systematic reviews to identify, appraise and synthesize the available evidence on cancer, cognitive impairment, adverse birth and pregnancy outcomes, oxidative stress, electromagnetic hypersensitivity, and heat-related effects.

The current manuscript describes the protocol that will be used to conduct the systematic review on carcinogenic risks of experimental RF EMF exposure in laboratory animals.

1.2. Description of the outcome

The health outcome in relation to RF EMF exposure for which the evidence will be assessed systematically is cancer risk in experimental laboratory animals.

Animal cancer bioassays have a long history of identifying the carcinogenic potential of environmental factors including chemical, physical and biological agents, and are considered a reliable approach to hazard identification of potential human carcinogens (IARC, 2019a) (https://doi.org/10.1002/9780470025079.chap26.pub2) and are also widely used for cancer risk assessment, particularly if human data do not allow a quantitative dose-response analysis. The term animal cancer bioassay as used in this protocol includes short- and long-term carcinogenicity studies, including life-time studies with laboratory animals exposed to RF EMF, initiation-promotion studies, co-carcinogenesis studies, both being models where cancer is induced by a substance used as initiator or promoter, and RF EMF exposure, studies in tumor-sensitive transgenic animals, and implantation (cancer cells) studies by all routes of exposure (Ashby, 2001; Boorman et al., 1994; IARC, 2019a; National Toxicology, 1976; National Toxicology Program, 2015a, b; OECD, 2018; Pastoor and Stevens, 2005). Medium-term duration tests for carcinogenicity on the development of proliferative lesions in a single tissue, e.g., foci of alteration in the liver, are also considered (OECD, 2018).

Relevant endpoints for cancer outcomes in animals have been defined by IARC and include tumor incidence and prevalence, tumor type, different stages of carcinogenesis including preneoplastic lesions, number of benign and malignant tumors per animal, and survival of the animals (IARC, 2019a). These endpoints depend on the experimental model (long-term cancer bioassay over about 2 years in rats and mice, initiation-promotion and tumor co-carcinogenesis studies, transgenic animal models, implanted tumor cells), species and strain as well as sex and tumor type. In initiation-promotion and tumor-co-carcinogenesis studies, the agent used for initiation and promotion drives the tumor type, and effects of RF EMF on those tumor types can be evaluated. Similar, in animal models with implanted tumor cells, only that tumor type can be compared to the long-term animal cancer bioassay, effects of RF EMF on tumor types in all organs and tissues / body fluids can be studied.

1.3. Description of the exposure

RF EMF are defined as fields with frequencies from 100 kHz to 300 GHz. Human exposure to RF EMF may occur in the environment and in the workplace from sources being either close to or far from the body, resulting in localized near-field and whole-body far-field exposure conditions, respectively. Near-field refers to distances from the RF EMF source less than a few wavelengths, for example, approximately 1 m at 1 GHz, whereas far-field behaviors dominate at greater distances.

Sources of exposure in occupational settings may be RF polyvinyl chloride welding machines and radar systems, while magnetic resonance imaging, a widely used diagnostic practice, may represent a source of exposure to medical personnel as well as to patients. The use of wearables using WLAN has increased. Mobile phones, desktop computers, laptops and tablets used with WLAN are the main sources of near-field exposure and common sources of far-field exposure include mobile phone base stations, radio and television antennas, digital enhanced cordless telecommunication (DECT) base stations, wireless local area network (WLAN, WiFi) access points, baby monitors, and smart meters. Future wireless communication will rely on current emerging technologies in wireless communication such as 5/6G and Internet of Things (IoT) solutions (https://standards.ieee.org/standard/1528_7-2020.html)

In addition to the distance from the field, the main variables influencing the interaction of RF EMF with the human body are the signal frequency (the higher the frequency, the lower the penetration depth), the exposure level (defined as the strength of the incident electric and magnetic fields) and the exposure duration, but also polarization of the field emitted by the source, modulation of the signal and dielectric characteristics of tissues play a role.

While exposure describes the electromagnetic fields (EMF) from the sources at a location where an individual might be present. Exposure to RF EMF results in induction of internal electric fields. The most relevant exposure metric (below 6 GHz) is the Specific Absorption Rate (SAR, W/kg tissue weight). The SAR (or 'dose') depends on the RF signal characteristics as frequency, external field strength, polarization and modulation, and on the characteristics of the absorbing tissue (dielectric parameters) and there is a direct proportionality with the (square) internal electric field.

In near-field conditions, the unique exposure metric is the SAR (W/kg), which can be measured with different techniques including computed relationships between SAR and the incident fields or direct measurement of temperature rise. Hence, in animal studies, the reported SAR includes values estimated and averaged over specified time and volume. Two metrics, in particular, are most often determined, namely the whole-body SAR and a time-averaged organ- or tissue-specific SAR.

When the exposure takes place in far field conditions, the internal field (and hence SAR) can be derived (deduced) from the measurement of the external electrical field strength (Volt per meter, V/m), since quantitative relationships between them can be established via sophisticated computational models. In fact, ICNIRP suggests limits for both internal ("basic restrictions") in terms of SAR and, because of the technical difficulty of evaluating internal dose, as external quantities for electric- or magnetic- field strength measures or as power density (W/m^2)("reference levels").

Exposure does not only depend on the source emissions and the geometry relative to the source, but also on the effect of the electromagnetic environment on the fields' incident. Hence, in experimental animal studies, processes such as shielding, reflection and diffraction need to be controlled and described properly. Furthermore, the distribution of the induced fields also depends on parameters of the source (e.g. frequency,

polarization), distance and location of the source with respect to the body, anatomy, dielectric properties, body posture and the environment. Hence, depending on the exposure of the animals in different exposure setups (carrousel versus cage, for example), these parameters need to be considered.

At high frequencies, above 6 GHz, the absorption is limited to the surface of the exposed target and the Absorbed Power Density (expressed in W/m^2) can be used as an appropriate tissue internal exposure metric.

Depending on the exposure level and time, energy absorption can result in temperature increase, thermoregulatory responses or changes in energy balance in the exposed target, which may induce heat-related biological effects. This temperature increase can either be directly measured using non-perturbing probes or predicted, using sophisticated modelling techniques.

1.4. Rationale for a systematic review

With the introduction of technologies based on RF EMF and the widespread use of mobile phones and other wireless devices, public concern has been raised as to whether exposure to RF EMF associated with these new technologies might be carcinogenic to humans. Experimental animal and epidemiological research have been conducted to investigate if RF EMF exposure is carcinogenic. In a first evaluation of the potential carcinogenicity, in 2011 IARC evaluated RF EMF as possibly carcinogenic to humans, based on limited evidence of carcinogenicity in humans and in animals (IARC, 2013). Since this evaluation, many new studies involving the three major evidence streams (cancer in humans, cancer in experimental animals, and mechanistic data) have been published and a re-evaluation has also been identified as a high priority by an international Advisory Group to the IARC Monographs program (IARC, 2019a,b).

2. Objectives and PECO statement

The question if RF EMF exposure leads to an increased risk of cancer in laboratory animals will be addressed systematically assessing eligible studies on animal cancer bioassays.

The evaluation of cancer in animals is based on the guidelines from IARC, OECD and the National Toxicology Program (NTP) (IARC, 2019a; National Toxicology Program, 2018; OECD, 2012; 2014; 2018) as well as the OHAT guidelines (National Toxicology Program, 2015b; National Toxicology Program, 2019a; National Toxicology Program, 2019b) and the Report on Carcinogens (RoC) (National Toxicology Program, 2015a), and the evaluation of the overall certainty of the evidence from animals studies will be developed following the Grading of Recommendations Assessment, Development and Evaluation (GRADE) framework.

A Population, Exposure, Comparator, Outcome (PECO) statatement was developed to focus the research question(s), identify search terms, and develop inclusion/exclusion criteria used during screening. The PECO statement was based on addressing the following questions:

(1) Among laboratory animals, does exposure to RF EMF increase the risk of cancer compared to sham, no exposure or lower exposure in experimental studies, and to assess if there is an exposure-dependent and/or time-dependent relation between the exposure and the outcome?
(2) Is there an exposure/dose response between the exposure and the outcome?

3. Methods

This systematic review will be carried out following the recommendations for systematic reviews by WHO adapted for laboratory animal studies (Dishaw et al., 2020; Hooijmans and Ritskes-Hoitinga, 2013; IARC, 2019a; Morgan et al., 2016; National Toxicology Program, 2015a; National Toxicology Program, 2019a; Rooney et al., 2014; Whaley et al., 2020; WHO, 2014). Data will be reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Moher et al., 2009; Moher et al., 2015). Tools and criteria for data extraction, Risk of Bias (RoB) assessment and rating of the overall evidence will be adapted to the specificity of animal carcinogenicity studies according the guidelines of the National Toxicology Program/Office of Health Assessment and Translation (NTP/OHAT) tool (National Toxicology Program, 2015a; b; National Toxicology Program, 2019a; National Toxicology Program, 2019b).

DistillerSR (https://www.evidencepartners.com/products/distillersr-systematic-review-software/) and Health Assessment Workplace Collaborative (HAWC; (https://hawcproject.org) will be used for literature priorization, data recording and sharing, data extraction and analyses (Shapiro et al., 2018).

3.1. Eligibility criteria

3.1.1. Types of populations

We will include studies conducted with non-human mammalian animal species (whole organism), of any life-stage (including preconception, *in utero*, lactation, peripubertal, and adult stages), of any strain, substrain and sex including transgenic animals. Non-human mammalian model systems (typically rodents) are widely used in toxicology and preferred over other model systems where greater uncertainty exists on predictivity for identifying human health hazards, e.g., fish, *C. elegans*, insects, etc. (IARC, 2019b; National Toxicology Program (NTP), 2015; National Toxicology Program, 2019b), IRIS Handbook:
https://ofmpub.epa.gov/eims/eimscomm.getfile?p_download_id=541571.

3.1.2. Types of exposures

3.1.2.1. RF EMF. We will include studies that have applied electric, magnetic or electromagnetic fields in the frequency range of 100 kHz to 300 GHz and that reported exposure using at least one of the situations listed below. These different situations are to accommodate different exposure metrics and exposure types reported in studies, allowing an optimal degree of inclusivity whilst at the same time ensuring confidence in there being a specified contrast between experimental and control conditions. These situations are as follows:

(A) ***body/tissue/sample internal exposure metrics*** *measured or calculated for the particular conditions of the experiment,*
- *SAR* (*expressed in* W/kg *or equivalent units*)
- *SA* (*expressed in* J/kg *or equivalent units*)
- *induced electric field strength, (expressed in V/m or equivalent units)*
- *internal magnetic field strength (expressed in A/m or equivalent units)*

  (for exposure applied as pure or predominantly magnetic fields in the lower frequency range *the external magnetic field strength at sample position* is considered a sufficient surrogate for the *tissue internal magnetic field* as long as the penetration depth is high compared to the sample dimension)

(B) ***body/tissue/sample internal exposure metrics*** *describing superficial absorption at frequencies above 6 GHz measured or calculated for the conditions of the experiment as follows,*
- *incident power flux density* (expressed in W/m^2 or equivalent units)
- *incident energy density* (expressed in J/m^2 or equivalent units)
- *transmitted (absorbed) power flux density* (expressed in W/m^2 or equivalent units)

- *transmitted (absorbed) energy density* (expressed in J/m^2 or equivalent units),

(C) ***body/tissue/sample external exposure metrics***

- *external electric field strength* (V/m) (E > 1 V/m or E>$\sqrt{10}$*background level in unshielded environment, otherwise no restriction)
- *external magnetic field strength* (mA/m) (H > 2.7 mA/m or H>$\sqrt{10}$*background level in unshielded environment, otherwise no restriction)
- *incident power flux density*, (mW/m^2 (PD > 2.5 mW/ m^2 or PD > 10*background level in unshielded environment, otherwise no restriction)

We will only include studies reporting external metrics as under C if (i) either of these exposure metrics was measured or calculated at the location of the exposed body in the approximate far-field of the field source, and (ii) the exposure level is at least a factor of 10 (power flux density) or $\sqrt{10}$ (field strength) above background level[1]. In the case where no specific background exposure level in the laboratory is reported in the study, we will assume a value of 0.25 mW/m^2 (corresponding to 0.3 V/m and 0.9 mA/m, respectively) as the background exposure level. This results in an inclusion threshold of PD = 2.5 mW/m^2, E = 1 V/m, or H = 2.7 mA/m).

(D) ***mobile phones or other RF-generating devices as source of exposure without reporting of metrics under A, B or C***

We will consider these studies separately because there is likely a greater variation and/or uncertainty in exposure levels owing to these being, in most cases, inferred rather than directly measured.

(1) ***with output controlled by appropriate software or hardware operated close to the tissue/sample***

We will include studies that applied exposure with an output power controlled by hardware or software, provided that the output power and the distance to the sample are reported, enabling inference of the exposure.

(2) ***in GSM mode with an active call operated close to the body/tissue/sample***

An exposure applied as the field generated by a mobile phone in GSM mode with an active call operated at distances equal or less than 3 cm from the body/tissue/sample can be expected to generate a temporal peak SAR in the range of 0.01 to 100 W/kg, that means at least a factor 100 above the average background level. We will include these studies only if the active call was maintained throughout the experiment and the comparison was a similar phone switched off.

It is noted that in operating modes other than GSM (e.g., UMTS, LTE), the power control of mobile phones is substantially more efficient than in GSM mode, resulting in the possibility of transmit power levels which are too low to ensure sufficient exposure contrast above background level. Therefore, only mobile phones in GSM mode with an active call operated close to the body/tissue/sample are considered to ensure sufficient exposure contrast.

We will include studies with an exposure duration of > 10 days for long-term animal cancer bioassays or at least 1 day in a transgenic rodent assay with at least one dose-level tested (OECD, 2009).

We will exclude studies that:

- have applied exposure signals with more than 10% of the total signal energy outside the considered frequency range, 100 kHz – 300 GHz (e.g., pulsed fields, non-sinusoidal fields with dominant frequencies <100 kHz).
- have applied exposure with a mobile phone in cases where at least one of the following conditions apply: i) the mobile phone was not operated in GSM mode and the output power was not controlled by hardware or software specifically for the experiment; ii) no active call was established and maintained during the experiment, because of the potentially extremely small exposure contrast generated which we do not consider relevant.

3.1.2.2. Co-exposure. Studies involving other exposures, e.g. in animal models where a tumor initiator and/or tumor promotor is used, will be included only if they include an experimental arm with exposure to RF EMF only. Studies involving co-exposure to a tumor initiator, tumor promotor or co-promotor need to provide the dose and timing of the administration, as well as the source and purity of the agent.

3.1.3. Types of comparators

We will include studies that have compared exposure to a concurrent control, namely a sham-exposed group or a group that has been exposed to a substantially lower level of RF EMF. (i.e. that are subjected to the same handling, laboratory environment, diet, and treatments as the RF-EMF exposed animals, with the exception of actual RF-EMF exposure) or a non-exposed control group (cage control with the laboratory environment e.g., humidity, noise may vary compared to the sham- and RF-EMF-exposed animals). For studies comprising more than one exposure level, the effects reported for each level of exposure will also be used to calculate dose–response relationships.

We will also include studies with historic controls (Dinse and

[1] Data from the literature (Jalilian et al., 2019) suggest RF E-fields of approximately 0.1–0.3 V/m (corresponding to PD of 25–250 μW/m^2) as a reasonable range for typical indoor RF background exposure in a laboratory environment. From other studies (Christ et al., 2006a, 2006b; Schmid and Kuster, 2015), it can be derived that such background exposure levels may lead to SAR values inside superficial tissue layers and cell cultures in the order of several tens to several hundreds of μW/kg, and that a RF source with a low directivity antenna (e.g. mobile phone) operated within a few centimeters of the body or biological sample may cause SAR values in superficial layers in the range of several tens to lower hundreds W/kg per watt transmit power. Hence, a mobile phone in GSM operating mode with an active call (peak transmit power in the range approx. 1 mW up to 2W) in close proximity to the body/tissue/sample is expected to cause substantially higher (approx. factor 100–1000) temporal peak SAR than typical RF background exposure, even when the transmit power of the mobile phone (in GSM mode) was not controlled by the experiment, but autonomously controlled by the network. Despite the expected high variation of exposure in such cases, we will therefore also include studies that used a mobile phone in GSM mode with an active call at distances equal or less than 3 cm from the body/tissue/sample as field source, provided that evidence is given that the active call was maintained throughout the experiment and that the control group (sham exposed group) was treated with the mobile phone switched off.*References for the footnote:*Christ, A.; Samaras, T.; Klingenbock, A.; Kuster, N. Characterization of the electromagnetic near-field absorption in layered biological tissue in the frequency range from 30 MHz to 6,000 MHz. Phys Med Biol 2006;51:4951–4965Christ, A., Klingenbock, A., Samaras, T., Goiceanu, C., Kuster, N., 2006a. The dependence of electromagnetic far-field absorption on body tissue composition in the frequency range from 300 MHz to 6 GHz. IEEE Transactions on Microwave Theory and Techniques 54, 2188–2195. https://doi.org/10.1109/TMTT.2006.872789Christ, A., Samaras, T., Klingenböck, A., Kuster, N., 2006b. Characterization of the electromagnetic near-field absorption in layered biological tissue in the frequency range from 30 MHz to 6000 MHz. Phys. Med. Biol. 51, 4951–4965. https://doi.org/10.1088/0031-9155/51/19/014Jalilian, H., Eeftens, M., Ziaei, M., Röösli, M., 2019. Public exposure to radiofrequency electromagnetic fields in everyday microenvironments: An updated systematic review for Europe. Environmental Research 176, 108517. https://doi.org/10.1016/j.envres.2019.05.048Schmid, G., Kuster, N., 2015. The discrepancy between maximum in vitro exposure levels and realistic conservative exposure levels of mobile phones operating at 900/1800 MHz. Bioelectromagnetics 36, 133–148. https://doi.org/10.1002/bem.21895

Peddada, 2011; IARC, 2019a; Tarone, 1982) because they provide sufficiently valid results if they fulfil the following conditions: 1. same sex, 2. same species, 3. same strain, 4. same gender, 5. same diet, and 6. same laboratory environment, 7. species that originates from the same laboratory within a time frame of not more than 5 years (Haseman, 1995). Less weight will be given to historical controls when they show a high degree of variability, and greater weight when they show a low variability. Historical controls are selected to resemble the concurrent controls as closely as possible with respect to species, strain, gender, route of exposure, diet and general laboratory environment (Greim et al., 2003; Haseman et al., 1984). We use historical controls for rare tumors ($\leq$1–3%) or when an abnormal control incidence response (e.g., a viral infection in control animals or other issues as identified in the ROB and Sensitivity tool) is seen (Dinse and Peddada, 2011; Kobayashi and Inoue, 1994). An abnormal control incidence response is a control value below normal historical control range or above the normal historical control range compared to a historical database in the facility and the breeding company.

3.1.4. Types of outcomes

We will include any study that has evaluated the incidence of one or more of the following cancer-related endpoints in any organ, tissue or body fluids of the laboratory animals:

(1) malignant tumours,
(2) pre-neoplastic lesions
(3) benign tumors
taking into account tumor latency, severity, and multiplicity of neoplasms and pre-neoplastic lesions.
(4) Combinations of benign and malignant tumors (a) they occur together with and originate from the same cell type as malignant tumors in an organ or tissue in a particular study and (b) they appear to represent a stage in the progression to malignancy (Huff et al., 1989).

3.1.5. Types of studies

3.1.5.1. Inclusion criteria. We will include experimental cancer studies performed in a controlled environment with laboratory animals randomly or non-randomly assigned to the exposure categories. Good Laboratory Practices (GLP) and Non-GLP studies will be included. Quality of an animal study should be considered, as to use of SPF-animals, lack of infectious diseases, appropriate numbers of animals per experimental group, proper experimental monitoring, and approved relevant toxicology and pathology protocols, etc.

We will include long-term carcinogenicity studies, initiation-promotion studies, co-carcinogenesis studies, studies in tumor-sensitive transgenic animals, and implantation (cancer cells) studies by all routes of exposure (OECD, 2018). Medium-term duration tests for carcinogenicity on the development of proliferative lesions in a single tissue, e.g., foci of alteration in the liver, will also be considered (OECD, 2018).

3.1.5.2. Exclusion criteria. Studies with exposure of laboratory animals outside the considered frequency range, 100 kHz – 300 GHz will be excluded as well as studies without data on exposure metrics and studies without sham-exposed animals or other controls (e.g., cage) with the RF EMF exposure in the controls being substantially lower than the RF EMF exposure.

3.1.5.3. Years considered. No year-of-publication restriction will be applied.

3.1.5.4. Publication language. No language restriction will be applied. Articles in languages other than the ones spoken by the reviewers (English, French, Italian, German, Japanese) will be translated into English using Deepl (https://www.deepl.com). If necessary, help from native speakers or interpreters will be sought. However, considering that title and abstract of non-English articles published in peer-reviewed journals are in English, only English terms will be used to search the publication databases.

3.1.5.5. Publication types. Original data from peer-reviewed publications and from reports and databases publicly available from government agencies will be considered. We will exclude pre-prints, abstracts, proceedings and case reports.

3.1.6. Types of effect measures

Eligible studies have to provide an effect measure or sufficient data to calculate an effect estimate; the type of effect measure will not influence the decision to include or exclude studies. For dichotomous outcomes, the absolute risk (AR) compared to suitable controls is used as the measure of the effect. If data are available for more than two exposure levels, we will calculate incremental ARs per unit of exposure increase.

The tumor incidence data have to be corrected for survival time of the experimental animals per group if survival time is different between RF EMF-exposed animals and concurrent controls, further the lethality of the tumor, and the number and size of tumors may require consideration (IARC, 2019a).

In order to compare the tumor incidence data across studies with different durations of exposure but the same sex, species and strain, adjustments may need to be made. There are two components of the study design that must be addressed, exposure level and time on study. Concentration (exposure level or dose) is typically applied in a randomized manner and is fixed. Time on study, however varies (because of biologic and stochastic variability between animals and/or the design of the study). Survival differences within studies will be corrected as needed as discussed earlier. Duration differences that are part of the study design are typically addressed by applying Haber's rule as modified by Tenberge et al. (1986). This rule assumes that $C^nT = k$ where C is the exposure in appropriate units, T is the duration of exposure, n is a constant derived from experimental data or from historical information and k is an unknown constant. In most cases, regulatory authorities use $n = 3$ (TCEQ, 2015) as a standard and that will be done here. Thus, for two studies with different durations of exposure (T_1 and T_2) and two different levels of exposure (C_1 and C_2) in the same units, it is assumed that:

$$C_1^3 T_1 = k = C_2^3 T_2 \text{ or}$$

$$C_1^3 T_1 = C_2^3 T_2 \text{ or}$$

$$C_1 = C_2 \left(\frac{T_2}{T_1}\right)^{1/3}$$

3.2. Information source and search strategy

The following publication databases will be used to search for eligible studies: MEDLINE via PubMed (https://pubmed.ncbi.nlm.nih.gov/), Science Citation Index Expanded and Emerging Sources Citation Index via Web of Science (https://clarivate.com/webofsciencegroup/solutions/web-of-science/), and the EMF Portal (https://www.emf-portal.org/). The complete literature search strategy is provided in Fig. S1 Annex 1.

The search strategy is designed via text analysis of key studies identified by the team. Search terms consist of controlled vocabulary (e.g. MeSH) and free text for the following main concepts: electromagnetic fields, cancer, and animal studies. An animal studies search filter by The National Toxicology Program (NTP) is adapted and applied to the search strategies.

Search filter: https://ntp.niehs.nih.gov/ntp/roc/handbook/rochandbookappendix_508.pdf

Search terms that specify the exposure (e.g., microwaves, high frequency fields, radio waves, radio frequency fields) or relevant exposure sources (e.g., mobile phones, GSM, UMTS) are used. No publication date limits will be applied.

The outcome of the searches is tested and if necessary, these and other terms are chosen by expert judgment through an iterative trial-and-error process. The goal is to obtain a list of articles as inclusive as possible.

The EMF Portal database (https://www.emf-portal.org/en) will be searched through the hierarchical organization of its domains: for instance, starting from 'Mobile communications', which automatically excludes the low frequency (50/60 Hz) EMF studies and studies on static fields. The focus of this search is on experimental studies in laboratory animals and cancer. The search query will contain specific keywords based on our PECO queries.

Records retrieved from the literature searches will be imported to Endnote® 20 for deduplication. The remaining records will be imported to DistillerSR for screening (https://www.evidencepartners.com/products/distillersr-systematic-review-software/). These records will be screened for eligibility in sequence by title and abstract, and then full text review. We will check the references lists of all eligible publications to identify studies not captured by the database searches. Forward citation searches will be performed via the Scopus database. An electronic library of included publications will be created in the Endnote® format.

We will rerun the literature searches to identify more recent studies before the final evaluation.

3.3. Study selection

Studies identified from the database searches will be imported into DistillerSR software to assess if they don't fulfill the inclusion criteria. Both title/abstract (TIAB) and full-text screening is conducted by two independent reviewers and all references chosen as relevant by a reviewer will be included in the full-text assessment. During TIAB screening, only studies that are clearly irrelevant will be excluded. Any study possibly fulfilling inclusion criteria will be screened at the full-text level. For citations with no abstract, articles are initially screened based on all, or some of the following: title relevance (title should indicate clear relevance). Conference reports and editorials will be excluded. Eligibility status of non–English studies will be assessed using the same approach with online translation tools or engagement with a native speaker used to facilitate screening. Full-text records are sought for studies screened as meeting PECO criteria or "unclear" based on the TIAB screening. References that are not able to be procured within 45 days of attempt will be determined to be unavailable.

The screening decisions are then imported into EPA's version of Health Assessment Workspace Collaborative (HAWC, https://hawcprd.epa.gov/portal/), a free and open source web–based software application designed to manage and facilitate the process of conducting literature assessments. In HAWC, the screening and tagging results are visualized in interactive literature tag trees where additional tagging can be conducted, e.g., more details on the type of animal cancer bioassay and the type of exposure.

HAWC (https://hawcproject.org/) and DistillerSR will be used to document the selection of eligible studies. In case of agreement between the reviewers, the article will be automatically moved to the full text phase. In case of disagreement, a discussion will follow to solve the issue. If no consensus can be reached, a third reviewer will be consulted. Study evaluation results, including the justifications for reviewer judgements, will be documented. The selection process is illustrated in a study flow diagram according to the PRISMA reporting guidelines (Page et al., 2021a; Page et al., 2021b).

3.4. Data extraction

Data extraction will be performed in HAWC (https://hawcproject.org/), and the tool will be adapted to animal cancer bioassays and RF EMF exposure. Accurate and reliable data collection is achieved by a customized list of data items (Annex 2) using the templates form from HAWC. Pilot-testing facilitates reliable data collection (WHO, 2014).

When there are multiple publications using the same or overlapping data, all publications will be included, with one selected for use as the primary study; the others will be considered as secondary publications with annotation in HAWC indicating their relationship to the primary record during data extraction.

Data from relevant studies will be extracted by one reviewer and each record will be verified by a second reviewer. Discrepancies during data extraction will be initially discussed by extractors and another team member will be involved if no agreement is reached. When there are multiple publications using the same or overlapping data, all publications will be included, with one selected for use as the primary study; the others will be considered as secondary publications with annotation in HAWC indicating their relationship to the primary record during data extraction. For animal studies, the primary publication will typically be the one with the longest duration of exposure, or with the outcome(s) most informative to the PECO. For corrections, retractions, and other companion documents to the included publications, a similar approach to annotation will be taken and the most recently published data will be incorporated in the assessments.

The characteristics and the outcome results of the studies will be documented in HAWC including information on study identification and methods (study characteristics such as, study design), animal model, strain, sub-strain and sex, exposure, and comparator(s), outcome(s), and statistical methods, and others (e.g., author queries) (National Toxicology Program, 2019a). Data on cancer outcomes (e.g., number of animals with tumor, system, tissue or organ name, neoplastic or non-neoplastic, name of tumor-based on histopathology / dose / group) will be extracted for each endpoint. For dichotomous variables, values necessary to calculate the effect measures will be extracted (e.g., number of events and number of animals/replicates in either the experimental or control group); for continuous outcomes, number of animals and means or medians with standard deviations/errors in each experimental group will be extracted.

Outcome data can be downloaded from HAWC as Excel files that can directly be used for further analyses (e.g., *meta*-analysis).

We will ask authors for missing data, particularly that involving missing key reporting quality information or other data important for conducting a *meta*-analysis or additional analyses that could address major study limitations identified during risk of bias assessment. Outreach to study authors is documented in HAWC and considered unsuccessful if researchers do not respond to an email or phone request within one month of the attempt to contact, and a reminder has been sent.

3.5. Risk of bias and sensitivity assessment

RoB assessment will be conducted using the NTP OHAT RoB tool (National Toxicology Program, 2015b; National Toxicology Program, 2019a; National Toxicology Program, 2019b) with input from EPA/IRIS (https://hawcproject.org/), and implemented in EPA's publically available version of HAWC (https://hawcproject.org/). Since the NTP/OHAT guidelines are not specifically designed for animal cancer bioassays, adaptations, especially the inclusion of sensitivity questions, were made using NTP's sister program, namely the Report on Carcinogens (RoC) methodology (National Toxicology Program, 2015a).

Key concerns are potential sources of bias (factors that could systematically affect the magnitude or direction of an effect) or insensitivity (factors that limit the ability of a study to detect a true effect). In brief, study evaluation judgements are reached for the following risk of bias

domains: 1. Selection bias, 2. Performance bias, 3. Detection bias, 4. Attrition/Exclusion bias, 5. Selective reporting, 6. Other sources of bias (see Annex 3). In addition, a sensitivity domain has been added. All of the OHAT, RoC, and EPA IRIS study evaluation tools probe the concept of sensitivity, typically in the exposure assessment and outcome assessment domains. Recognizing the concepts of risk of bias and sensitivity are distinct, we have created a separate domain for sensitivity to enhance transparency (for details see RoB Annex 3) and described the tool as a RoB and sensitivity tool. The sensitivity assessment will be conducted separately from the RoB. Signaling / prompting questions will be used to guide the judgement for each domain and are described in more detail in the NTP RoC Handbook (National Toxicology Program, 2015a). After a pilot phase to calibrate judgements, each study will be independently reviewed by two reviewers to reach consensus ratings at the domain level. Based on OHAT guidelines (EPA/IRIS guidelines), one of four risk of bias scores will be assigned to each bias question: 'definitely low risk of bias' (good), 'probably low risk of bias' (adequate), 'probably high risk of bias' (deficient), and 'definitely high risk of bias' (critically deficient) (National Toxicology Program, 2019a). Instructions for response are provided in the detailed guide for using the tool Sensitivity judgements will be characterized as "no/minor concerns" or "critical concerns (National Toxicology Program, 2019a): at least two reviewers independently evaluate each study, after a pilot phase is conducted, to calibrate judgements across the team. The independent reviewers use structured web-forms for study evaluation housed within the EPA's publicly available version of HAWC (https://hawcproject.org/) to record separate judgments for each domain and the overall study for each outcome, to reach consensus between reviewers, and when necessary, resolve differences by discussion between the reviewers or consultation with additional reviewers. Explanatory guidelines on how to assign scores on the basis of a predefined set of criteria are given in the structured web-forms in HAWC. Results are displayed using a 'heatmap' format. The rationale for the classification, including a quote from the article, and a brief description of any identified strengths and/or limitations from the domains and their potential impact on the overall confidence determination is documented and retrievable in HAWC (citation HAWC url for our project (SR2) will be included before publication). Briefly, quotes from the article will be added to the evaluation in HAWC, explaining the judgement score and the final decision, e.g., quote: 'animals were randomly assigned', and judgement: unclear how randomization was performed, decision: probably low risk of bias.

3.5.1. Excluding or analyzing studies based on aspects of risk of bias and sensitivity

We will use the tiering approach outlined by OHAT that favors inclusion of studies unless they are problematic in multiple RoB domains (National Toxicology Program, 2019b). According to the OHAT approach, studies are categorized as Tier 1, Tier 2 or Tier 3 based on RoB domain judgemens for RoB and sensitivity where Tier 3 studies are considered to have significant limitations. Typically, Tier 3 studies have limitations in multiple RoB domains. In the current analysis, study sensitivity will also be considered as a non-RoB domain to inform tiering judgements. These tiering judgements will be used to inform subsequent analyses (National Toxicology Program, 2019a). Specifically, the tiering categories will be used as a basis to inform sensitivity and stratified analyses to assess whether RoB/sensitivity issues were contributing to heterogeneity in the evidence base and robustness of results.

3.6. Synthesis of results

We will perform a random effects *meta*-analysis of the RRs of the highest exposure contrasts in studies with similar PECOs (see section on heterogeneity for what is sufficiently similar to be combined). When data are available for more than two exposure levels, we will calculate incremental risks via *meta*-regression. We will combine the results of studies considered to be similar in narrative way if a *meta*-analysis is not possible. A narrative synthesis of the tabulated results will be provided regarding the questions based on PECO. When a *meta*-analysis is not possible, the similarity of the effect sizes will be judged regarding the direction of the outcome (e.g., increased or decreased risk of cancer for a respective tumor type).

Only studies that are considered sufficiently similar will be combined. Studies with laboratory animals of the same species, strain, sex and cancer outcome will be combined if their study design (animal cancer bioassay type, and lengths of the study, e.g., 2-year cancer bioassay) is sufficiently similar.

As defined above, relevant endpoints for cancer outcomes include tumor incidence and prevalence, tumor type, different stages of carcinogenesis including preneoplastic lesions, number of benign and malignant tumors per animal, and survival of the animals (IARC, 2019a; b). These endpoints depend on the experimental model (long-term bioassay over about 2 years, initiation-promotion and tumor co-carcinogenesis studies, transgenic animal models, implanted tumor cells), species and strain as well as sex and tumor type.

If we find an increased risk of cancer, we will also evaluate the following sub-questions:

(1) For which tumor types is the risk increased?
(2) Is there a difference in effect between species- strain- or sex?

3.6.1. Assessment of heterogeneity

First, studies with any frequencies and types of exposure (near-field and far-field) will be combined. Second, studies with near-field RF EMF exposure are considered different from studies with far-field exposure, and the same applies for different modulations (pulsed exposure).

Statistical heterogeneity of results will be quantified by the I^2 and the τ^2 statistics (Higgins and Thompson, 2002; Higgins et al., 2003; Higgins et al., 2009), taking into account the methodological limitations of these test statistics. I^2 presents the inconsistency between the study results and quantifies the proportion of observed dispersion that is real, i.e. due to between study differences. The heterogeneity measure, I^2 reflects the extent of overlap of the confidence intervals of the study effects. I^2 represents the inconsistency on a scale between 0 and 100, and therefore, it can be compared with suggested limits for low or high inconsistency. The τ (the square root of τ^2) is the standard deviation of the between-study variation on the scale of the original outcome. The τ^2 is the direct estimate of the between-study variation, and it can be used to estimate prediction intervals for the combined effect size measure calculations. In addition to the I^2 and the τ^2 statistics, an 80% prediction interval will be estimated from the distribution of effect estimates being the interval of effect estimates comprising the true effect size for 80% of analyzed studies (IntHout et al., 2016). However, a *meta*-analysis is only performed when at least 2 studies are considered similar enough to be combined (https://cccrg.cochrane.org/author-resources). However, certain statistical tests typically used in *meta*-analyses (e.g., test for heterogeneity or publication bias) require larger number of studies. If the number of studies is too low, potential heterogeneity will be assessed by visual inspection of the respective results from individual studies.

3.7. Additional analyses

3.7.1. Subgroup analyses

When possible, any metrics related to cancer risk will be quantified and a narrative synthesis or a *meta*-analysis using appropriate subgroups will be chosen, with the latter including studies of the same exposure type.

Subgroup analyses will be performed for similar exposure by frequency and pattern of exposure (continuous or pulsed) and outcome(s) specified in the PECOs. Only studies considered sufficiently homogenous, namely same species, strain, sex, tumor type, and frequency and pattern of exposure will be combined.

If we find an increased risk of cancer (malignant tumors), benign tumors only or a combination of benign and malignant tumors, we will also evaluate the following sub-questions:

(1) Is there a difference in effect between exposure with high versus low frequencies of RF EMF?
(2) Is there a difference regarding exposure type (near-field versus far-field)?
(3) Is there a difference between pattern of exposure (continuous versus pulsed exposure)?

3.7.2. Sensitivity analyses

The impact of any potential RoB or concerns for study insensitivity will be assessed by comparing results of the studies that are at low RoB or no/minor concerns noted for sensitivity with the results of studies with high RoB or critical concerns noted for sensitivity for a specific domain and evaluate if these are statistically different. This would entail an initial comparison of studies considered Tier 1 or Tier 2 to those indicated as Tier 3. Depending on this analysis, a more granular comparison of Tier 1 versus Tier 2 may be warranted. In addition, the sensitivity of assumptions made to assess exposure levels and categories will be evaluated.

3.7.3. Publication bias

First, a visual inspection will be performed to evaluate if small studies with no effect on cancer are missing. If these seem to be missing we will apply Egger's linear regression analysis to evaluate potential small study bias if 10 or more studies are included in the same *meta*-analysis (Egger et al., 1997).

3.8. Certainty of evidence assessment

We will apply GRADE as described in the OHAT Handbook (National Toxicology Program, 2019a) adapted to the subject matter on RF-EMF exposure and cancer in experimental animals applied to animal studies to qualify the certainty of the evidence for each conclusion on the effect of the exposure on an outcome for each category/frequency of exposure.

In order to judge the certainty in the evidence of the effects observed in the SRs and draw reliable conclusions, a framework based on the GRADE principles will be used (https://www.gradeworkinggroup.org) as outlined in Annex 4. Similar to recommendations for human experimental studies, the rating on the certainty of evidence for animal studies starts at high certainty evidence (Hooijmans et al., 2018). Depending on the overall evaluation of a study, the certainty of evidence is downgraded to moderate, low or very low.

Reasons for downgrading are: limitations in studies as indicated by the RoB across studies, indirectness, inconsistency, and imprecision. Indirectness for animal cancer bioassays is evaluated by assessment if the PECO has been addressed appropriately.

Five properties for a body of evidence (large magnitude of effect, dose response, residual confounding, and consistency across study designs and experimental model systems, other, such as specificity of the association in cases where the effect is rare) will be used to determine if the initial confidence rating should be upgraded (National Toxicology Program, 2019a).

The GRADE findings will be contextualized in the narrative to convey conclusions regarding direction and magnitude of effect(s) (Hultcrantz et al., 2017; Santesso et al., 2020).

Financial support

This project is partly funded by the World Health Organization.

Author contributions

Starting from WHO problem formulation and protocol drafts, MM and KS led the team and together with KT coordinated protocol development including the tools for data extraction, RoB and GRADE. TR developed the search strategy based on the input provided by MM. KT and KS provided expertise on systematic review methodology and risk of bias assessment. Expertise on RF EMF dosimetry and exposure assessment was provided by AWW. Expertise on animal carcinogenesis outcome assessments in experimental animals and cancer bioassays, was provided by JW, JM, MM, KS and AKS. Expertise on statistical evaluations was provided by AKS. All authors contributed to editing and finalizing the protocol.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: AWW directs a research group, which includes three technical associates who are telecommunications company employees. JM receives employment and research support from The government of Canada related to the topic. KS has been the Head of the IARC Monographs program until his regular retirement (11/2018). Since 10/2019, he is a member of the International Scientific Advisory Committee of the Ramazzini Institute. This involves one 3 h advisory group meeting per year. He does not receive remuneration for his advisory activity. MM is a member of the scientific advisory board of The Swiss Research Foundation for Electricity and Mobile Communication (FSM) that receives research money from commercial entities. She does not receive remuneration for his advisory activity. Her partner does consulting relating to cell phone safety. All remaining authors declare no conflict of interest.

Acknowledgments

We greatly acknowledge the support from Emilie van Deventer, Jos Verbeek, Eric van Rongen, and the principal investigators of the other SR teams. We thank Shoji Fukushima for his support.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.envint.2022.107106.

References

Ashby, J., 2001. Expectations for transgenic rodent cancer bioassay models. Toxicol. Pathol. 29 (1_suppl), 177–182.
Boorman, G.A., Maronpot, R.R., Eustis, S.L., 1994. Rodent carcinogenicity bioassay: past, present, and future. Toxicol. Pathol. 22 (2), 105–111.
Dinse, G.E., Peddada, S.D., 2011. Comparing tumor rates in current and historical control groups in rodent cancer bioassays. Stat. Biopharm. Res. 3 (1), 97–105.
Dishaw, L., Yost, E., Arzuaga, X., Luke, A., Kraft, A., Walker, T., Thayer, K., 2020. A novel study evaluation strategy in the systematic review of animal toxicology studies for human health assessments of environmental chemicals. Environ. Int. 141, 105736. https://doi.org/10.1016/j.envint.2020.105736.
Egger, M., Smith, G.D., Schneider, M., Minder, C., 1997. Bias in meta-analysis detected by a simple, graphical test. BMJ 315 (7109), 629–634.
Greim, H., Gelbke, H.-P., Reuter, U., Thielmann, H.W., Edler, L., 2003. Evaluation of historical control data in carcinogenicity studies. Hum. Exp. Toxicol. 22 (10), 541–549.
Haseman, J.K., 1995. Data analysis: statistical analysis and use of historical control data. Regul. Toxicol. Pharmacol. 21 (1), 52–59.
Haseman, J.K., Huff, J., Boorman, G.A., 1984. Use of historical control data in carcinogenicity studies in rodents. Toxicol. Pathol. 12 (2), 126–135.
Higgins, J.P.T., Thompson, S.G., 2002. Quantifying heterogeneity in a meta-analysis. Stat. Med. 21 (11), 1539–1558.
Higgins, J.P., Thompson, S.G., Deeks, J.J., Altman, D.G., 2003. Measuring inconsistency in meta-analyses. BMJ 327, 557–560.
Higgins, J.P.T., Thompson, S.G., Spiegelhalter, D.J., 2009. A re-evaluation of random-effects meta-analysis. J. Roy. Stat. Soc. Ser. A Stat. Soc. 172 (1), 137–159.
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., Minnerup, J., 2018. Facilitating healthcare decisions by assessing the certainty in the evidence from preclinical animal studies. PLoS One 13 (1), e0187271.

Hooijmans, C.R., Ritskes-Hoitinga, M., 2013. Progress in using systematic reviews of animal studies to improve translational research. PLoS Med. 10 (7), e1001482. https://doi.org/10.1371/journal.pmed.1001482.

Huff, J.E., Eustis, S.L., Haseman, J.K., 1989. Occurrence and relevance of chemically induced benign neoplasms in long-term carcinogenicity studies. Cancer Metastasis Rev. 8 (1), 1–21.

Hultcrantz, M., Rind, D., Akl, E.A., Treweek, S., Mustafa, R.A., Iorio, A., Alper, B.S., Meerpohl, J.J., Murad, M.H., Ansari, M.T., Katikireddi, S.V., Östlund, P., Tranæus, S., Christensen, R., Gartlehner, G., Brozek, J., Izcovich, A., Schünemann, H., Guyatt, G., 2017. The GRADE Working Group clarifies the construct of certainty of evidence. J. Clin. Epidemiol. 87, 4–13.

IARC. Non-Ionizing Radiation, Part 2: Radiofrequency Electromagnetic Fields. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans, Volume 102.1 eds: IARC Press; 2013.

IARC, 2019a.. Preamble to the IARC Monographs. In: Humans I.m.o.t.e.o.c.r.t., ed. Lyon, France.

IARC, 2019b. Tumour Site Concordance and Mechanisms of Carcinogenesis ed. Lyon, France: International Agency for Research on Cancer.

IntHout, J., Ioannidis, J.P.A., Rovers, M.M., Goeman, J.J., 2016. Plea for routinely presenting prediction intervals in meta-analysis. BMJ Open 6 (7), e010247. https://doi.org/10.1136/bmjopen-2015-010247.

Kobayashi, K., Inoue, H., 1994. Statistical analytical methods for comparing the incidence of tumors to the historical control data. J. Toxicol. Sci. 19 (1), 1–6.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G., 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. J. Clin. Epidemiol. 62 (10), 1006–1012.

Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P., Stewart, L.A., 2015. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst. Rev. 4 (1) https://doi.org/10.1186/2046-4053-4-1.

Morgan, R.L., Thayer, K.A., Bero, L., Bruce, N., Falck-Ytter, Y., Ghersi, D., Guyatt, G., Hooijmans, C., Langendam, M., Mandrioli, D., Mustafa, R.A., Rehfuess, E.A., Rooney, A.A., Shea, B., Silbergeld, E.K., Sutton, P., Wolfe, M.S., Woodruff, T.J., Verbeek, J.H., Holloway, A.C., Santesso, N., Schünemann, H.J., 2016. GRADE: Assessing the quality of evidence in environmental and occupational health. Environ. Int. 92-93, 611–616.

National Toxicology, P., 1976. Guidelines for carcinogen bioassay in small rodents. Natl. Cancer Inst. Carcinog. Tech. Rep. Ser. 1, 1–65.

National Toxicology Program (NTP), 2015. U.S.D.o.H.a.H.S. Handbook for Preparing Report on Carcinogens Monographs ed˜eds: Office of the Report on Carcinogens Division of the National Toxicology Program National Institute of Environmental Health Sciences U.S. Department of Health and Human Service.

National Toxicology Program, N., 2015a. Handbook for Preparing Report on Carcinogens Monographs. Office of the Report on Carcinogensm Divition of the National Toxicology Program, National Institute of Environmental Health Sciences, U.S. Department of Health and Human Services.

National Toxicology Program, N., 2015b. OHAT Risk of Bias Rating Tool for Human and Animal Studies. Office of Health Assessment and Translation. In: Science O.o.H.A.a.T.N.I.o.E.H., ed.

National Toxicology Program, N., 2018. Toxicology and carcinogenesis studies in Hsd: Sprague Dawley sd rats exposed to whole-body radio frequency radiation at a frequency (900 MHz) and modulations (GSM and CDMA) used by cell phones. Natl. Toxicol. Program Tech. Rept. Series 1–384.

National Toxicology Program, N., 2019a. Handbook for Conducting a Literature-based Health Assessment Using OHAT Approach for Systematic Review and Evidence Integration. In: Sciences. O.o.H.A.a.T.N.I.o.E.H., ed.

National Toxicology Program, N., 2019b. OHAT Risk of Bias Rating Tool for Human and Animal Studies. In: NIEHS/NTP, ed.

OECD, 2009. Detailed review paper on transgenic rodent mutation assays, series on testing and 9 transgenic rodent gene mutation assay. OECD, Paris, France.

OECD, 2012. Guidance Document 116 on the Conduct and Design of Chronic Toxicity and Carcinogenicity Studies. In: Environment Health and Safety Publications, ed. Paris: OECD.

OECD, 2014. Guidance Document 116 on the Conduct and Design of Chronic Toxicity and Carcinogenicity Studies, Supporting Test Guidelines 451, 452 and 453 ed or eds.

OECD, 2018. Test No. 451: Carcinogenicity Studies ed or eds.

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., Hrobjartsson, 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., 2021a. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. Int. J. Surg. 88, 105906.

Page, M.J., Moher, D., 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., Hrobjartsson, 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., McKenzie, J.E., 2021b. PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. BMJ 372, n160.

Pastoor, T., Stevens, J., 2005. Historical perspective of the cancer bioassay. Scand J. Work Environ. Health 31(Suppl 1) 129–140; discussion 119-122.

Rooney, A.A., Boyles, A.L., Wolfe, M.S., Bucher, J.R., Thayer, K.A., 2014. Systematic review and evidence integration for literature-based environmental health science assessments. Environ. Health Perspect. 122 (7), 711–718.

Santesso, N., Glenton, C., Dahm, P., Garner, P., Akl, E.A., Alper, B., Brignardello-Petersen, R., Carrasco-Labra, A., De Beer, H., Hultcrantz, M., Kuijpers, T., Meerpohl, J., Morgan, R., Mustafa, R., Skoetz, N., Sultan, S., Wiysonge, C., Guyatt, G., Schünemann, H.J., 2020. GRADE guidelines 26: informative statements to communicate the findings of systematic reviews of interventions. J. Clin. Epidemiol. 119, 126–135.

Shapiro, A.J., Antoni, S., Guyton, K.Z., Lunn, R.M., Loomis, D., Rusyn, I., Jahnke, G.D., Schwingl, P.J., Mehta, S.S., Addington, J., Guha, N., 2018. Software tools to facilitate systematic review used for cancer hazard identification. Environ. Health Perspect. 126 (10), 104501. https://doi.org/10.1289/EHP4224.

Tarone, R.E., 1982. The use of historical control information in testing for a trend in proportions. Biometrics 38 (1), 215. https://doi.org/10.2307/2530304.

TCEQ, 2015. Guidelines to Develop Effects Screening Levels, Reference Values, and Unit Risk Factors RG-442. Texas Commission on Environmental Quality.

Tenberge, W.F., Zwart, A., Appelman, L.M., 1986. Concentration-time mortality response relationship of irritant and systemically acting vapors and gases. J. Hazard. Mater. 13, 301–309.

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.

Whaley, P., Aiassa, E., Beausoleil, C., Beronius, A., Bilotta, G., Boobis, A., de Vries, R., Hanberg, A., Hoffmann, S., Hunt, N., Kwiatkowski, C.F., Lam, J., Lipworth, S., Martin, O., Randall, N., Rhomberg, L., Rooney, A.A., Schünemann, H.J., Wikoff, D., Wolffe, T., Halsall, C., 2020. Recommendations for the conduct of systematic reviews in toxicology and environmental health research (COSTER). Environ. Int. 143, 105926. https://doi.org/10.1016/j.envint.2020.105926.

WHO, 2014. WHO Handbook for Guideline Development, 2nd. ed. WHO Press, World Health Organization, Geneva.