

How to deal with dissent among experts. Risk evaluation of EMF in a scientific dialogue

Holger Schütz and Peter M. Wiedemann

Forschungszentrum Jülich GmbH
Programmgruppe Mensch, Umwelt, Technik
52425 Jülich
Germany

Corresponding author:

Holger Schütz
Forschungszentrum Jülich GmbH
Programmgruppe Mensch, Umwelt, Technik
52425 Jülich, Germany
Tel.: +49 2461 61-3595
Fax: +49 2461 61-8175
Email: h.schuetz@fz-juelich.de

Authors' Note: We are grateful to two anonymous reviewers who commented on a first version of this article. The research project was supported by a grant from T-Mobile Germany.

Abstract

This paper describes the results of a scientific discourse which aimed at exploring the reasons for differences in expert health risk assessment of radio frequency electromagnetic fields of mobile telephony.

The paper starts with describing the structure of the discourse. Then, the reasons for the conflicting risk assessments are discussed. Differences are due to the selection and evaluation of relevant scientific studies by applying different scientific quality standards, to the methods used for generating a research synthesis and an overall risk evaluation. Consensus could be achieved regarding the selection of and the quality requirements for the scientific studies used for risk assessment as well as their significance for risk evaluation. However, dissent remained about the synthesis of scientific evidence into an overall risk evaluation and about the relevance of the precautionary principle for risk evaluation and its implications for the risk assessment framework.

Based on the analysis of these problems, a transparent, consistent and rational procedure for risk assessment is suggested to facilitate a risk characterization which better meets the demands of policy making and the public for an appropriate risk evaluation.

Keywords: Risks of mobile telephony, scientific dialogue, evaluation of evidence, risk assessment, risk characterization

Introduction

Expert disagreements are an important reason for public conflict about a risk issue, and it is evident that they also pose a severe problem for risk communication. A typical example are the potential health risks of radio frequency electromagnetic fields (RF EMF) of mobile telephony which have raised public concerns all over Europe and are an ongoing topic in public debate in many countries (Burgess, 2004). The main issue is whether the existing exposure limits – which have been suggested by the International Commission on Non-Ionizing Radiation Protection (ICNIRP, 1998) – are sufficient to protect public health.

These exposure limits are based on the scientific evidence about thermal (i.e. heating) effects of EMF exposure. Whether exposure below these exposure limits might cause detrimental health effects, is scientifically controversial. As a result, although many countries adopted the ICNIRP recommendations for exposure limits, other countries – for instance Switzerland, Poland or Luxembourg – implemented stricter precautionary limit values.

Over the last years, a number of international expert reports have evaluated the scientific evidence on potential health risks from EMF exposure below the ICNIRP exposure limits (e.g. BUWAL, 2003; Health Council of the Netherlands, 2000, IEGMP, 2000, Royal Society of Canada, 1999, SSK, 2001, Zmirou, 2001). All of these reports agree that there is no scientific proof of health risks below these exposure limits. However, the reports show disagreements with regard to the extent and the relevance of uncertainties in the scientific knowledge about this issue, and in particular they disagree whether precautionary measures should be implemented or not.

Expert workshops and consensus conferences have been suggested as a means to deal with such disagreements and uncertainties (e.g. OECD, 2002). They focus on the formation of a commonly accepted body of data, analyses and models, i.e. they aim to support scientific understanding and to narrow uncertainty and thereby are putting gentle pressure on participating scientists to find a consensus. One may hope that in a critical scientific dialogue, expert disagreements can be specified and discussed and, perhaps, can be resolved. But even if no agreement can be achieved, the reasons underlying the disagreements should become clear, and this should provide a useful input for risk management.

The present paper develops a proposal for structuring and conducting such a scientific dialogue. This proposal is based on experiences drawn from a scientific dialogue on risk evaluation of potential risks of mobile telephony electromagnetic fields. In the first part of the paper, the background, structure and problems of that dialogue are reported. The second part outlines the structure of a scientific dialogue which tries to overcome these problems and provides a transparent procedure for establishing a risk assessment when expert disagreements exist.

I. Experiences from a scientific dialogue on potential RF EMF risks

Background

An opportunity for setting up an expert dialogue arose when in late 1999 the German mobile phone operator T-Mobile commissioned four German scientific institutions to prepare its own expert opinions on possible health risks from mobile telephony electromagnetic fields. The contractors of these studies were scientists from the Ecolog Institute, Hannover (expert A), the Institute for Applied Ecology, Darmstadt/Freiburg (expert B), the Humboldt University, Berlin (expert C), and from the RWTH Aachen University of Technology (expert D). They represent a range of “scientific orientations” in the EMF debate from “critical” (experts A and B) to “established” (experts C and D) science. For each the task was to select the most important 100 scientific studies, to evaluate these studies with regard to the evidence they provide about possible health risks from radio frequency electromagnetic fields and to summarize the evaluations in a risk assessment. The expert opinions were completed between April and November 2000.

The Programme Group Humans, Environment, Technology (MUT) of Research Centre Jülich started a dialogue with these experts in July 2001, which was completed in September 2002. The aim of this dialogue was to render the consensus and dissent between the expert opinions available and the reasons for these. It was intended to verify in a fair debate whether the arguments of the respective risk assessments are transparent and scientifically reliable. The aim was thus to make transparent how the different conclusions drawn by the experts have been brought about. Wiedemann et al (2003) give a detailed account on the scientific dialogue.¹

Structure of the dialogue

Based on standards for evidence assessment and risk characterisation which have been set by the International Commission on Non-Ionizing Radiation Protection (ICNIRP, 2002), the American Environmental Protection Agency (EPA, 2000), the Australian National Health and Medical Research Council (NHMRC, 2000) and the International Agency for Research on Cancer (IARC, 1999), MUT developed a structure for the scientific dialogue. This dialogue comprised six steps.

The first step was to check the agreement in the scientific studies selected by the four experts as a data base for their reports. Due to different priority setting and search strategies, considerable differences may arise in the selection of the

¹ The project report and the four expert opinions are also available (in German) on the Internet: www.emf-risiko.de

studies. If such differences in the data base exist, differences in risk assessment are not surprising.

As the expert opinions covered a wide range of endpoints relevant for assessing the effects on health produced by the electromagnetic fields of mobile telephony, in the second step six research fields, which were considered to be of particular importance, were selected together with the four experts in a first workshop. These research fields were discussed in detail in the following workshops (see Table 1).

In the third step, for each of the selected six research fields a consulting expert was incorporated who was familiar with the methodology of the respective research field. His task was not to provide an assessment of health risks from EMF, but to evaluate the reasonableness of the four experts' assessments from a methodological perspective and to answer specific technical questions which came up during the workshop discussions. This turned out to be an important feature of the process, since the range of relevant topics (cancer, reproduction disturbances, cognitive performance, stress hormones, chromosomal mutations, cell communication etc.) was so large that it could no longer be competently assessed by an individual in all details. In addition, the interpretation of subtle effects sometimes requires a person's own experimental experience with the investigation methods applied. However, scientists only have such experience in their own fields of research. They must therefore rely on their colleagues' support in other fields of research.²

In the fourth step the explanatory power of the scientific studies were examined and a scientific overall picture was established. Guidance for this step was be drawn from the approach of evidence-based medicine, which provides rules on how to proceed in examining scientific investigations to establish a scientific overall picture. Questions for these examinations include: Can chance results be excluded? Have sampling errors and other errors been minimized? Have confounding variables been eliminated or at least been controlled?

In the fifth step a final characterization of the pros and cons regarding the existence of a health risk from RF EMF was produced. That is, arguments speaking for and against such a risk are elaborated and weighed against each other (see also EPA, 2000). This step was guided by the risk evaluation guidelines of the Californian EMF project (Neutra et al 2002). The most important criteria for weighing the arguments are the power of the study design to explore a causal relationship between EMF exposure and effect as well as the relevance of the study result to the assessment of health impacts on humans. This characterization of the pros and cons was carried out as an iterative process, which started with the pros and cons as they were expressed in the expert opinions and during the workshops. These were then elaborated and condensed by MUT.

² In selecting consulting experts we tried to ensure that these scientists not only exhibit high expertise for the respective research field, but also that they did not represent a particular faction in the debate on the EMF health risks.

As the sixth step, the pro and con arguments were discussed on a final workshop with the experts and consulting experts, and changed or extended where necessary. Finally, the four experts – each from his perspective – weighted these pros and cons once again.

The scientific dialogue was in a series of workshops (see Table 1). The three workshops, in which the specific research fields were discussed in detail, had the same overall structure. In the beginning MUT presented the results of the four expert assessments for the selected research fields. Afterwards, the four experts were invited to comment this presentation and to clarify their risk assessments. Next, the consulting expert presented his view on the usefulness of data and methods in his area of expertise for risk assessment and he evaluated the four expert opinions.

- Table 1 about here -

Checking the data base of the expert opinions

The starting point for the analysis of the data base was the reference lists contained in the four expert opinions. In total, 320 different studies rated as important and a total of 483 studies are referred to in the four expert opinions. However, there is little agreement among the experts with respect to the literature referred to.

Only 5 % or 16 of the 320 important studies are found in all four expert opinions. 73 % or 232 of the 320 studies are only taken into account by one expert, 13 % are referred to in two of the four expert opinions and 9 % in three of the expert opinions. Moreover, the agreement in the topics selected for the dialogue project varies considerably: it ranges between 0 % and 50 % agreement for all four expert opinions (see Figure 1).

- Figure 1 about here -

There are a number of reasons for these pronounced differences in study selection. It may be assumed, for example, that some investigations were published several times in different journals. In this case, the relevant criterion for assessing the agreement in the data base would not be the incorporation of a specific publication, but rather of a specific investigation result. Another reason is that experts have sometimes refrained from citing older studies of an author or a group of authors, if in their more recent studies the earlier results are also presented.

However, the differences in the data base also result from the experts' diverse selection criteria. This manifests itself particularly drastically for the topic of epidemiology, where two experts only take account of the two epidemiological studies relating directly to mobile phones then available. On the other hand, the two other experts also include studies in which the occupational exposure to RF EMF and the exposure to the fields of radio and TV stations are examined.

Other examples of different selection criteria are that expert B also discussed some studies on low-frequency fields (50/60 Hz), whereas the other experts do not take such studies into consideration, or that expert D is the only one who considered the influence of mobile phone fields on electronic implants.

As has been mentioned above the experts had been selected to reflect different orientations in science: Experts A and B represent the “critical” scientists, while experts C and D represent the “established” science. A rather obvious hypothesis is that these science orientations may play a role in the selection of studies. Thus, one could expect two “citation camps”, one formed by experts A and B, the other by experts C and D. However, it becomes clear from the figures in Table 2, that this hypothesis is not confirmed by the data. The cells in Table 2 show the intersections between each two expert opinions. For instance, the “critical” expert A has a larger intersection with the “established” expert D than with the other “critical” expert B. And although the “established” expert C cites more studies which are also cited by the other “established” expert D, the latter in turn has more studies in common with experts A and B than with expert C. Thus, the figures suggest that there are no “citation camps” of experts based in their science orientation.

- Table 2 about here -

Our analysis also showed that the differences in the data base are only in a few cases due to the fact that certain studies were overlooked by individual experts. Nevertheless, the finding remains that the experts make a different selection. This points to the necessity of making selection criteria explicit, of justifying and harmonizing them.

Although the low agreement of the expert opinions concerning the data base actually does not provide any information about the quality of the individual expert opinions, it nevertheless indicates a problem: It remains unclear to the reader whether differences in risk assessment result from the different data base or whether they are the result of other differences in risk assessment. And, of course, an obvious question is whether the differences in the data base might imply that essential scientific studies have not been taken into account in individual expert opinions and less important studies have been included – without the reader being able to decide which expert opinion is to be rated as better and which as less good.

It should be noted, however, that in spite of these differences in data base it was possible in the workshop discussions to agree on the studies essential for risk assessment for the respective research field.

Reasons for experts’ disagreement

The workshop discussions revealed a number of recurring problems, which are crucial for the differences in the expert opinions (see Table 3). These problems relate - beside the above mentioned differences in the data base underlying the

expert opinions - to (1) the scientific quality standards applied by the experts and their rating of the informativeness of the studies referred to, and (2) to for establishing of the scientific overall picture.

All these problems can be traced back, at least in part, to the different orientation – precautionary principle versus danger prevention – the four experts used as a basis for their evaluations and risk assessments. This issue will be discussed later in detail.

- Table 3 about here -

(1) Scientific quality standards

The experts basically agreed that the quality standards commonly used in science must be applied in evaluating the results of scientific investigations for risk assessment. Thus, for example, case study observations concerning the effect of mobile phone fields on humans cannot be referred to for risk assessment. It was also agreed that – measured against the stringent requirements, as formulated, for example, by the German Commission on Radiological Protection (e.g. replication required, unambiguous scientific overall picture; SSK, 2001) – the findings available cannot be regarded as proofs of damage to health. The experts disagreed, however, with respect to whether these findings should be regarded as an indication for risk. In the following, the critical issues underlying these differences will be described.

Sensitivity and specificity: For some of the issues concerning the effects of RF EMF investigated in the scientific literature, methods are used which respond very sensitively, but are not very specific – for instance the waking EEG and methods to investigate stress hormones and DNA strand breaks. Attention was drawn to these problems by the consulting experts for the respective topics. In the four expert opinions themselves, this aspect is not systematically and critically discussed as a problem in risk assessment. “Sensitivity” means that a method provides a small fraction of false negative findings. Associations that are present in reality are detected. A method is the more sensitive, the higher is the fraction of true positive findings. “Specificity” means that a method provides a small fraction of false positive findings. A method is the more specific, the higher is the fraction of true negative findings.

For the assessment of findings, high sensitivity but low specificity means that the effects found in experiments are not or only with difficulty causally attributable to EMF exposure. This implies that results obtained with the aid of such measurement procedures have no or only little informative value for risk assessment. In the discussions, the experts followed the judgement of the consulting experts, so that finally there was a consensus among the participants in the discussion that results from investigations in which the waking EEG, stress hormones or DNA strand breaks are examined have no or only little significance for the risk assessment of mobile phone fields.

Statistical significance and trends: In principle, there was agreement among the experts that the statistical support of study results is an important quality criterion for argumentation. However, at least some participants in the discussion took the view that statistical significance (1% or 5% significance level) should not be a "k.o." criterion in the evaluation of study results. This means that, when evaluating the significance of a result, not only the p-values, but also related aspects such as sample size and effect size – and thus the power of a study – or consistent, but non-significant trends in data should be considered. Notably one expert took the view that (statistically non-significant) trends, which become apparent in the data, also provide important information for the interpretation of the results – especially if these trends can be found in a similar way for different parameters.

Such a relaxation of the usual significance criterion implies that the probability of type I errors (i.e. false positive: an effect is accepted, although there actually is no effect) becomes greater and the probability of type II errors (i.e. false negative: an effect that is actually present is rejected) becomes smaller. Whether such a shift in criteria for acceptance or rejection of study results is justified in risk assessment especially from the aspect of precaution was a subject of a controversial debate, for which again no consensus could be reached.

Replication of studies: One of the central issues discussed was the necessity of replicating study results before these can be used for risk assessment. Although there was agreement among the participants in the discussion that replication is basically desirable and that such replications are lacking for many investigations on the effects of RF EMF, one expert emphasized that an exclusion of study results from risk assessment due to lacking replication would not be reasonable. It was argued here that in this field of research exact replications of investigations are often technically very difficult and sometimes even not appropriate (if, for example, flaws in the original study design are detected in retrospect).

Although these arguments were not challenged as possible explanations for the status quo it remained contentious among the workshop participants whether investigation results must have been replicated before they can be used for risk assessment. In this regard, it was pointed out that the claim for replication should not only apply to positive findings, but also for negative results. Another critical issue in the discussions was the question of how exactly a study must replicate the original in order to be considered as a replication. It was not possible to conclusively clarify this question either.

Relevance to health: There was agreement among the experts that a distinction must be made between the biological effects of RF EMF and damage to health due to such fields. This distinction is indeed central to risk assessment, since any effect is relevant for risk assessment only insofar as this effect can be considered to be a (potential) damage to health.

The experts did not basically rule out that RF EMF may have biological effects, however, they judged the reliability of the respective study results quite differently. More important, however, are the experts' different views concerning the question of whether and under what conditions biological effects can be used as indicators of adverse health effects in a risk assessment. For instance, is a shortening of response times under the influence of mobile phone fields, as found by Preece et al (1999), an indicator of health risks – although a shortening of response time is in the first place indicative of an improvement rather than a deterioration of cognitive functions?

One side argued that from the precaution perspective biological effects may also be important for risk assessment, even if they actually do not constitute any damage to health effects since they can point to action mechanisms of RF EMF as yet not understood. Although the other experts also noted deficiencies in scientific understanding for some biological effects, they did not follow the argument that biological effects per se can already be regarded as indicators of damage.

This different weighting of biological effects, on the one hand, and damage to health, on the other, leads to a different interpretation of the body of findings and thus also to different risk assessments.

Ecological validity: In addition to the problem of the relevance of findings to health, the ecological validity of the findings also played an important role in the discussions. The problem is whether (laboratory or epidemiological) studies with EMF which are in some regard different from EMF as emitted from mobile phones or mobile phone transmitters should be used for risk assessment of mobile telephony.

The experts agreed that not only studies performed in the mobile phone frequency range can be used for a risk assessment, but also those which investigated neighboring frequency ranges (100 MHz to 3 GHz). Thus, epidemiological studies with radio or TV stations as emission sources were judged to be informative by all participants in the discussion for the risk assessment of mobile phone fields. There was no basic dissent either concerning the significance of the signal shape (pulsed versus non-pulsed) for assessing the relevance of studies.

Different judgments were encountered, however, with respect to the requirements to be met by the exposure measurement (dosimetry). Not all experts took the view that findings from studies without a clear characterization of exposure are unsuitable for the risk assessment of mobile telephony. Especially one – he also was aware of this problem - but nevertheless included studies in his risk assessment, for which in most cases no sufficient exposure characterizations are made. He justified this by the fact that no better studies are currently available. Concerning this issue, the different views between the experts could not be resolved.

(2) Establishing the scientific overall assessment

In addition to the discussion of individual findings for the various endpoints, the expert opinions also provide a summary which tries to give an overall picture of the available evidence on possible effects of RF EMF on human health. Here the differences between the four expert opinions are conspicuous. There are examples of both the highlighting of positive or negative findings and the omission of findings. The question arises of how bodies of findings can be appropriately combined to obtain a scientific overall picture. Although this problem took up much less space in the workshop discussions than the problems of risk assessment so far described, its significance must not be underestimated.

Various methods for establishing a scientific overall picture are proposed in the literature (e.g. Cooper and Hedges, 1994). Whenever possible, a quantitative synthesis of the body of findings available should be made by means of a meta-analysis. However, a meta-analysis makes relatively high demands on the information available. Such a procedure is not found in any of the four expert opinions and would hardly have been possible for most of the topics due to the very heterogeneous body of data. In the expert opinions the findings are therefore summarized in a qualitative form, sometimes supported by tabular summaries of the individual studies. In establishing a scientific overall picture, a crucial issue is the extent to which findings should be weighted with respect to the quality of a study and other relevant criteria. Even more important, however, is the question of how to deal with different results of studies of comparable quality. Simple counting (“the majority of studies ...”) is not satisfactory here. It would rather be necessary to discuss the possible causes for the different findings in detail.

In the workshop discussions the crucial controversy concerned the special weighting of positive individual findings, as done by one expert. In the discussion, he defended this weighting with the argument that – from a precautionary perspective – these positive findings should be considered as valid as long as they are not refuted by means of an identically reproduced study.³ However, this view was not supported by the other experts, who insisted that positive and negative findings must be taken into consideration.

Results of the dialogue

The final weighing of the pro and con arguments, which was done by the four experts at the concluding workshop, revealed what was apparent already during the previous workshop discussions: Although several arguments, particularly from the consulting experts (e.g. about the appropriateness of waking EEG measurement for risk assessment), challenged a number of the assessments given by the experts, none of the experts was really willing to revise his risk evaluation. Therefore it was also not possible to arrive at an agreement between the experts about possible risks from RF EMF of mobile telephony to

³ Note that this argumentation is at odds with the critical view of this expert on replication as a practical scientific quality standard (see above).

human health. In this respect the dialogue process cannot be regarded as successful. However, the dialogue process was successful in disclosing the reasons for disagreements and thus made these disagreements comprehensible.

Furthermore it became evident that the evaluation of the existing evidence (i.e. the selection of studies, the application of scientific quality standards and the establishing of a scientific overall assessment) is based on elaborate interpretations and tacit knowledge. This leaves space for differing assessments and conclusions and allows to repel new arguments or evidence which challenge one's own judgements. Under these circumstances existing preassumptions (and their underlying values and attitudes) which guide the evaluation of evidence seem to be as important as the evidence itself.

Our results have implications for the design of further scientific dialogues. They suggest that rather than starting a dialogue process *after* different risk assessments have been given by experts, the scientific dialogue should start *before* expert opinions are commissioned. The dialogue process should be used as a tool for organizing and structuring the risk assessment process by different experts. This will be outlined in the following.

II. Proposals for a scientific dialogue on risk assessment

In addition to the experiences from the scientific dialogue, reported here, there are a number of proposals for a structured and transparent risk assessment (see also the survey in Wiedemann et al 2002). For example, the German Expert Council for Environmental Issues (SRU, 1996) submitted a procedure proposal developed by the German Risk Commission (Risikokommission, 2002). In the USA, proposals have been presented by the National Research Council (1996) and the Presidential/Congressional Commission on Risk Assessment and Risk Management (1997). However, these proposals have found only little resonance in the EMF-risk assessment community which is referring mainly to technical standards and less to the social side of "conducting science" (see e.g. Repacholi and Cardis, 1997). In the following we will outline a proposal for conducting a risk assessment which considers both sides – the essential criteria and standards for reviewing evidence as well as the design of a process which is fair and balanced and competent.

Composition of the risk assessment group (stage 1, see Figure 2): Whenever a risk assessment requires different scientific expertise, as it is the case in the field of mobile telephony EMF (dosimetry, biophysics, biology, sleep research, epidemiology, endocrinology, molecular medicine etc.), and certainly in many other risk fields as well, care must be taken to ensure that all essential fields of expertise are represented in the dialogue process.

Furthermore, the goals and frame of investigation (stage 2) must be determined at the beginning of the dialogue process. The central terms used for characterizing the body of findings must also be clarified. Agreement must be reached

among participants on the use of terms such as proof, suspicion, indication, adversity etc. Where possible, established definitions of recognized bodies (e.g. of the World Health Organization) should be used.

A standardized procedure must be developed for the selection of scientific studies (stage 3). Essential questions that must be clarified for this selection include:

- The type of studies to be taken into consideration: What quality requirements must be met by the studies? Should only studies be included that have been peer-reviewed, is so-called "grey literature" acceptable?
- What frequency domain is to be observed, what field intensities are relevant?
- How are the studies included to be documented?

Inclusion of consulting experts (stage 4) is an essential characteristic of the scientific dialogue. Their exact role and in particular the expertise needed must be determined. In addition to the obvious criterion of expertise in a specific field, two other criteria are important in selecting the consulting experts: they should be independent of economic or political interests in the mobile telephony debate and, if possible, they should not be involved in the scientific and/or public debate on possible risks of mobile telephony. In this way, it will be ensured that the consulting experts will not be accused of representing a particular faction, a frequent reproach in the public discussion of highly politicized topics, such as that of possible mobile telephony risks, in order to discredit disagreeable assessments.

In the analysis of evidence (stage 5) it should be clarified whether or not a study provides evidence for a causation of health-related effects on humans by the RF EMF of mobile telephony. For this assessment several criteria can be used (see Table 4).

- Table 4 about here -

Depending on the sensitivity and specificity of the investigation method, three directions of evidence are to be distinguished (Neutra et al 2002):

- Directionally neutral evidence: a positive result strengthens the evidence, a negative result weakens it. Example: well-designed animal cancer studies
- Predominantly weakening evidence: a negative result is more informative than a positive one, because the research method is sensitive, but not specific. Example: no effect is found in waking EEG examinations.
- Predominantly strengthening evidence: a positive result is more informative than a negative one, because the investigation is not very sensitive. Example: a positive result is found in an epidemiological investigation with short follow-up.

In order to establish the scientific overall picture (stage 6), it is essential that all the studies on the different endpoints and not just individual studies are taken into consideration. In this regard, the consistency of the body of findings, the uncertainties and the variability are to be discussed (see Table 5).

- Table 5 about here -

The risk characterization (stage 7) should provide an overall assessment of a risk, which is as complete and informative as possible and can be used for decision-making and which is based on and summarizes the preceding step of risk assessment (EPA, 2000, p. A-5). A risk characterization should include the following elements:

- The summary presents the central results as well as the essential strengths and weaknesses of the risk assessment and thus helps the risk manager to derive and decide on appropriate management options.
- The protection goals and targets of risk assessment must be explicitly specified. It must be made clear whether the aim is danger prevention or precaution. Finally, it must be clarified whether in addition to health and ecological risks also so-called "quality-of-life" risks are considered.
- The "context of risk assessment" provides information about the motive for risk assessment and about possibly already available assessments. Procedure specifications reveal whether and on what guidelines the risk assessment is based. It should also be specified here what consultations with what social groups have been performed for risk assessment. For better transparency of available expert knowledge, expert profiles should be generated and made accessible. Such profile discloses for each expert his or her scientific core expertise, the fields of professional work and the scientific bodies of which he or she is a member.
- Subsequently, the results of the individual steps of risk assessment must be described. Any quantitative criteria available to illustrate the risk or the cause for concern should be used (e.g. margin of exposure). If the risk is vague or can only be described qualitatively, the pros and cons of a suspected risk must be illustrated. In this case, a consistent definition and clarity of the assessment categories must be ensured for the understanding of the risk characterization.
- Classification of the findings in an evaluation scheme, e.g. as suggested by IARC (1999: probably carcinogenic, possibly carcinogenic, etc.) or SSK (2001: risk indication, suspected risk, proven risk).

These steps of a scientific dialogue are supposed to provide a structured process which allows for integrating different experts with possibly varying scientific orientations. Of course, adhering to this structure does not guarantee that in the end a consensus among the participants will be established. What should become clear, however, after such a scientific dialogue has been completed, is where the experts agree, where they disagree and what the reason are therefor. This should give decision makers a better orientation for their *risk evaluation* than a mere "coexistence" of different expert opinions for risk assessment.

In this risk evaluation, which is no longer a task of science alone because political valuations are incorporated here, are decisions on the guiding principles of evaluation – for instance, on the application of the precautionary principle – to be made. And only then can a decision be made concerning the selection of appropriate options for risk management.

Implications for further research

Future research should focus on the development of robust systems and rules for categorizing evidence. There are some good examples on which this research might build up. For instance, the IARC classification of carcinogenic substances (IARC 1999) also provides an appropriate starting point for the development of an evidence assessment scheme for other endpoints, such as somatic disturbances, changes on the molecular level, or disturbances of the blood brain barrier. Here, the problem of conflicting evidence or heterogeneity should be addressed with particular attention.

However, the most crucial issue in dealing with uncertain risks – for which RF EMF is a typical example – refers to the disclosure, characterization and communication of uncertainty and ignorance. Astonishingly, the empirical research is rather poor in this field and the effects of communicating uncertainties on risk perception, trust in experts, and appraisal of experts' competence are mixed. Therefore, it seems to be necessary that further research focuses on (1) the exploration of scientific hedging as a means to express uncertainties, (2) the assessment of the effects of formats of disclosure for uncertainties on risk understanding, and (3) the assessment of the influence of various devices to communicate different levels of uncertainty.

References

Burgess, A. (2004) *Cellular phones, public fears, and a culture of precaution*. Cambridge: Cambridge University Press.

BUWAL – Rööfli, M. and Rapp, R. (2003) *Hochfrequente Strahlung und Gesundheit* (BUWAL Umwelt-Materialien Nr. 162). Bern, Schweiz: Bundesamt für Umwelt, Wald und Landschaft (BUWAL).

Cooper, H. and Hedges, L.V. (1994) *The handbook of research synthesis*. New York: Russell Sage Foundation.

EPA (1999) *Guidelines for carcinogenic risk assessment*. U.S. Environmental Protection Agency, Washington, DC.

EPA (2000) *Risk characterization handbook*. United States Environmental Protection Agency. (on the Internet at: <http://www.epa.gov/osp/spc/rchandbk.pdf>), Washington, DC.

Health Council of the Netherlands (2000) *Mobile telephones. An evaluation of health effects*. The Hague: Health Council of the Netherlands, 2002, publication no. 2001/01E.

IARC (1999) *Preamble to the IARC Monographs*. IARC Monographs Programme on the Evaluation of Carcinogenic Risks to Humans. International Agency for Research on Cancer. (on the Internet at: <http://193.51.164.11/monoeval/Eval.html>).

ICNIRP (1998) Guidelines for Limiting Exposure to Time-Varying Electric, Magnetic, and Electromagnetic Fields (Up to 300 GHz). *Health Physics* 74(4), 494-522.

ICNIRP (2002) ICNIRP Statement: General Approach to Protection against Non-Ionizing Radiation. *Health Physics* 82, 540-548.

IEGMP – Independent Expert Group on Mobile Phones (2000) *Mobile Phones and Health*. National Radiological Protection Board. Chilton, UK.

Kheifets, L.I., Hester, G.L. and Banerjee, G.L. (2001) The precautionary principle and EMF: Implementation and evaluation. *Journal of Risk Research* 4, 113-125.

National Research Council (1996) *Understanding risk. Informing decisions in a democratic society*. Washington, DC: National Academy Press.

Neitzke, H.P. (2001) *Mobilfunk: Expositionen, Risiken, Vorsorge*. Ecolog-Institut Hannover. (on the Internet at: www.ecolog-institut.de/grenzwer.htm).

Neutra, R., DePizzo, V. and Lee, G.M. (2002). *An evaluation of the possible risks from electric and magnetic fields (EMFs) from power lines, internal wiring, electrical occupations and appliances* (Final Report). Oakland, CA: California EMF Project.
(Online unter: <http://www.dhs.ca.gov/ehib/emf/RiskEvaluation/riskeval.html>).

NHMRC (2000) *How to use the evidence: Assessment and application of scientific evidence*. National Health and Medical Research Council. (online available at: <http://www.nhmrc.health.gov.au/publications/pdf/cp69.pdf>).

Preece, A.W., Iwi, G., Davies-Smith, A., Wesnes, K., Butlers, S., Lim, E., Varey, A. (1999) Effect of a 915-MHz simulated mobile phone signal on cognitive function in man. *Int. J. of Radiation Biology* 75, 447-456.

Presidential Commission on Risk Assessment and Risk Management (1997) *Framework for Environmental Health Management. Volumes 1 & 2*. The Presidential/Congressional Commission on Risk Assessment and Risk Management. Washington, DC (on the Internet at: <http://cfpub.epa.gov/ncea/cfm/pcrarm.cfm?ActType=default>).

Repacholi, M.H. and Cardis, E. (1997) Criteria for EMF health risk assessment. *Radiation Protection Dosimetry*, 72, 302-312.

Risikokommission (2002) *1. Bericht der Risikokommission*. Geschäftsstelle der Risikokommission, Bundesamt für Strahlenschutz (Online at: <http://www.apug.de>).

Royal Society of Canada (1999) *A Review of the Potential Health Risks of Radiofrequency Fields from Wireless Telecommunication Devices*. An Expert Panel Report prepared at the request of The Royal Society of Canada for Health Canada. Ottawa, Ontario.

SRU (1996) *Der Rat von Sachverständigen für Umweltfragen: Umweltgutachten 1996. Zur Umsetzung einer dauerhaft umweltgerechten Entwicklung*. Metzler-Poeschel, Stuttgart.

SSK (2001) *Grenzwerte und Vorsorgemaßnahmen zum Schutz der Bevölkerung vor elektromagnetischen Feldern – Empfehlungen der Strahlenschutzkommission*. Bonn.: Strahlenschutzkommission.

WHO (2003) *Precautionary Framework for Public Health Protection* (Draft for Review, 2 May 2003). Geneva, Switzerland: World Health Organisation.

Wiedemann, P.M., Mertens, J., Clauberg, M., Hennings, W. and Schütz, H. (2002) *Umweltstandards*. Forschungszentrum Jülich GmbH. Programmgruppe Mensch, Umwelt, Technik.

Wiedemann, P.M., Schütz, H. and Thalmann, A.T. (2003) *Mobilfunk und Gesundheit* (Reihe Umwelt/Environment 42). Jülich: Schriften des Forschungszentrums Jülich.

Zmirou (2001) *Les Telephones Mobiles, Leurs Stations De Base et la Sante Etat des connaissances et recommandations*. Rapport au Directeur Général de la Santé.

Tables

Table 1: Topics addressed in the workshops

Workshop	Topic
Workshop 1	Selection of 6 research areas for closer discussion
Workshop 2	Cognitive processes
	EEG / awake and sleep
Workshop 3	Carcinogenic and teratogenic effects (animal studies)
	Carcinogenic and teratogenic effects (epidemiological studies)
Workshop 4	Calcium homeostasis, stress hormones
	Chromosomal mutations, DNA string breaks, cell proliferation
Workshop 5	Final evaluation and risk characterization

Table 2: Pairwise agreement in the selection of studies (absolute frequencies in parenthesis)

	Cited by expert A	Cited by expert B	Cited by expert C	Cited by expert D
Also cited by expert A	-	32% (36)	26% (32)	35% (43)*
Also cited by expert B	32% (36)	-	26% (32)	37% (46)
Also cited by expert C	28% (32)	29% (32)	-	30% (37)
Also cited by expert D	38% (43)*	41% (46)	31% (37)	-

* Read: 38 % of the references cited by expert A are also cited by expert D; 35% of the references cited by expert D are also cited by expert A. Differences in percentages with equal absolute frequencies result from the different number of references used by each expert.

Table 3: Problems of risk assessment

Topic	Problems
danger prevention vs. precaution	expert opinions proceed from different goals (danger prevention vs. precaution)
Selection / compilation of scientific literature / studies	there is little agreement among the expert opinions concerning literature selection
Sensitivity and specificity of the investigation methods	sensitivity and specificity are not treated systematically in the expert opinions
Statistical significance / trend	statistical significance is not considered necessary by all experts, if a trend can be observed
Replication of studies / reproduction of results	a replication is not considered necessary by all experts
Relevance to health	the relevance to health of biological effects is assessed differently
Ecological validity	there are no uniform requirements concerning the criteria required for the transferability of the studies to real situations
Summary and weighting of individual findings	different procedures are found for weighting the findings

Table 4: Criteria for evidence assessment for EMF health risks

Criterion	Lead question
Dosimetry	Are the relevant exposure parameters (field intensity, frequency domain, signal shape) of the investigation sufficiently known?
Statistical significance	Can the findings be sufficiently well safeguarded against the assumption of a random result?
Causality	Can other variables as an explanation of the association found be ruled out with sufficient certainty?
Adversity	Is the effect found of relevance to human health?
Ecological validity	Can the situation considered in the investigation be sufficiently well transferred to the real conditions under which mobile telephony takes place?

Table 5: Criteria for establishing the scientific overall picture

Criterion	Content
Consistency	The assessment of the body of findings in establishing the scientific overall picture provides important information for final risk characterization. This implies that both the positive and the negative findings (with their respective weighting) must be taken into account in establishing the scientific overall picture. Individual findings are to be weighted with a view to their methodological quality.
Uncertainties	It is necessary to make the uncertainties in exposure assessment explicit. Furthermore, it must be perceivable how certain or uncertain the knowledge about the adversity (i.e. the damaging effect) of the investigated end points is. Is it a scientifically uncontested fact, a hypothesis or a suspicion? The same also applies to the mechanism of action. Is this mechanism known, is there a scientific assumption or is no mechanism known? Moreover, it must also be discussed whether the findings from <i>in-vitro</i> studies and animal experiments can be transferred to humans. For this purpose, guidelines – such as those of the American Environmental Protection Agency (EPA 1999) for assessing carcinogenicity – can be used.
Variability	This means whether the findings, provided that they can be transferred to humans, take possible, different sensitivities – for example, of children – into account. However, it is important here that different dispositions are not only assumed, but also sufficiently confirmed. This also addresses the issue of electrosensitivity: it is still scientifically unclarified whether such a special disposition exists. These problems, which are ultimately again uncertainties, must be made explicit.

Figure

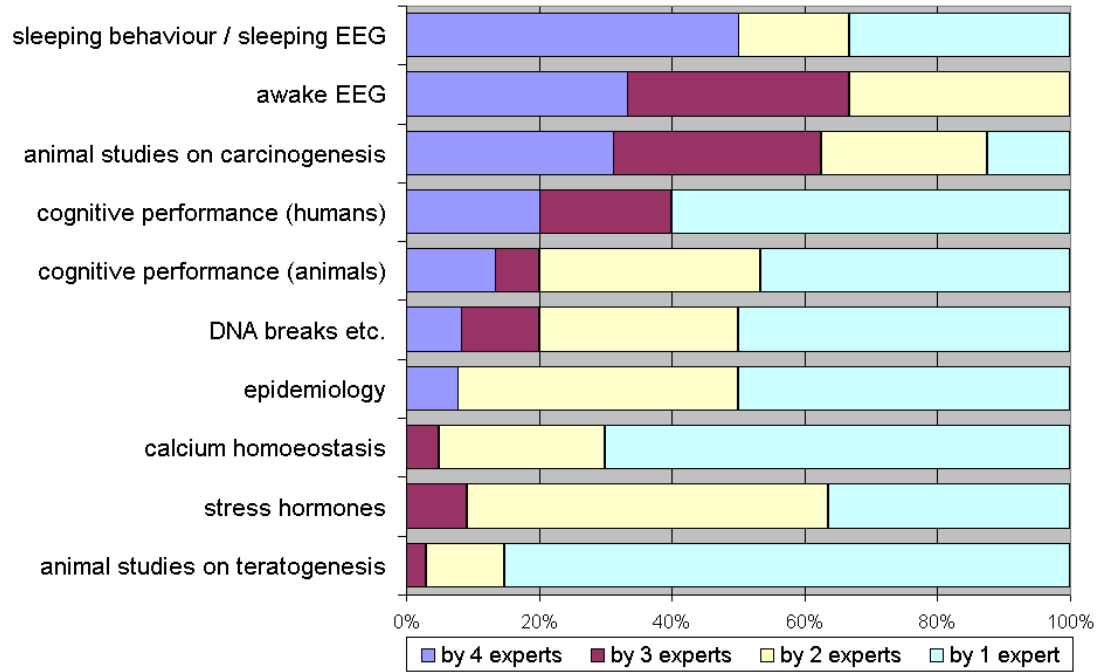


Figure 1: Percentage agreement of experts for the individual topics