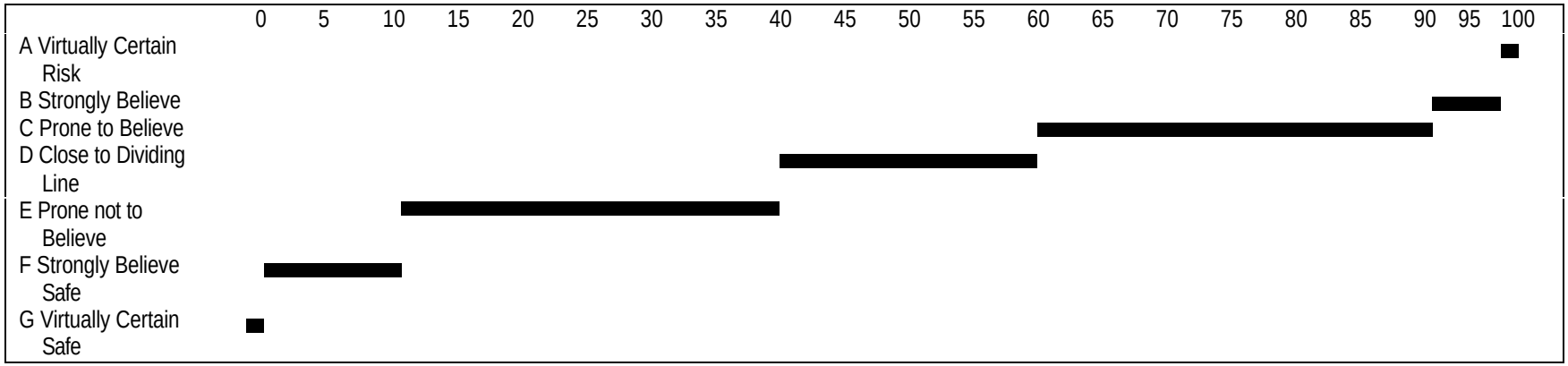
1 WHO DID THE EVALUATION AND WHAT FORM DID THE CONCLUSIONS TAKE?

On behalf of the California Public Utilities Commission (CPUC), three scientists who work for the California Department of Health Services (DHS) were asked to review the studies about possible health problems from electric and magnetic fields (EMFs) from power lines, wiring in buildings, some jobs, and appliances. The CPUC request for review did not include radio frequency EMFs from cell phones and radio towers. Reviewer 1, Vincent Delpizzo, Ph.D., is a physicist and epidemiologist; Reviewer 2, Raymond Richard Neutra, M.D., Dr.P.H., is a physician epidemiologist; and Reviewer 3, Geraldine Lee, Ph.D., is an epidemiologist with training in genetics. All three have published original research in the EMF area and have followed the field for many years. To integrate and extend their body of knowledge, the EMF Program contracted with specialists in biophysics, statistics, and animal experimentation to prepare a background in critical literature review in their respective fields and to make sure that the literature review was up to date through June 2000 (P. Gailey, Ph.D., G. Sherman, Ph.D., W. Rogers, Ph.D., and A. Martin, Ph.D.). The first three were involved with the writing of the 1998 National Institutes of Environmental Health Sciences (NIEHS) report. Furthermore, for each chapter of the review, another DHS epidemiologist or toxicologist was asked to read the original literature and consulted extensively with whichever of the three core reviewers was writing that chapter. This ensured that the writer based his/her evaluation on an understanding of the evidence that was as objective and consistent as possible. All three reviewers worked for the EMF program for at least five years and to some extent they influenced each other's thinking through their constant interaction and the review of each other's chapters. All three did their reviews according to the Risk Evaluation Guidelines (REG) that had been developed earlier and approved by the program's Science Advisory Panel (SAP). The Guidelines specified that the conclusions about any hazard should be done using two systems. The first was developed by the International Agency for Research on Cancer (IARC) and has been used by the NIEHS. It rates an agent as a Definite, Probable, Possible carcinogen or Not a carcinogen, or specifies that the evidence is "Inadequate" to rate the agent. In addition, the California Guidelines specified that in order to accommodate the probability-based computer models of the program's policy projects each of the DHS reviewers would individually assign a number between 0 and 100 to denote their degree of certainty that epidemiological associations between EMFs and certain diseases indicated that EMFs increased the risk of those diseases to some degree. They indicated their best judgement graphically with a little "x" and placed a shaded bar on either side of that "x" to indicate how uncertain they were. The best judgement and the uncertainty ranges could be used in quantitative policy analysis. The Guidelines, which were modified with advice from public comment and the SAP and the DHS reviewers, attached pre-agreed-upon English language phrases to various ranges of this degree of certainty. These are presented below in Table I.

If all three judges had best judgments above 50 out of 100, but that fell in different categories in Table I, judges were said to be "inclined to believe" that EMFs increased the risk of that disease to some degree.

TABLE I. EVERYDAY ENGLISH PHRASES TO DESCRIBE DEGREES OF CERTAINTY OF CAUSALITY (GRAPH ILLUSTRATES THE RANGE OF CERTAINTY NUMBERS TO WHICH THE PHRASES PERTAIN)

ARE THE HIGHEST EMFS AT HOME OR AT WORK SAFE, OR DO HIGH EMFS INCREASE THE RISK OF ..... TO A DEGREE DETECTABLE BY EPIDEMIOLOGY?	DEGREE OF CERTAINTY ON A SCALE OF 1 TO 100
Virtually certain that they increase the risk to some degree	>99.5
Strongly believe that they increase the risk to some degree	90 to 99.5
Prone to believe that they increase the risk to some degree	60 to 90
Close to the dividing line between believing or not believing that EMFs increase the risk to some degree	40 to 60
Prone to believe that they do not increase the risk to any degree	10 to 40
Strongly believe that they do not increase the risk to any degree	0.5 to 10
Virtually certain that they do not increase the risk to any degree	< 0.5



2 A SUMMARY OF WHAT HAS CHANGED SINCE THE CALIFORNIA EMF PROGRAM WAS FIRST PROPOSED IN THE EARLY 1990S

Between the time CPUC mandated a targeted California research program in 1993 to the time of this writing, considerable information has accumulated. In addition, three expert panels, the NIEHS Working Group (Portier & Wolfe, 1998), the IARC (IARC, 2001), and the British National Radiological Protection Board (NRPB, 2001b) have indicated that EMFs are a possible cause of childhood leukemia.

Biophysics: Biophysical arguments based on physical principles and simplified biological models have produced lower and lower predictions as to what magnetic field intensities theoretically would be capable of producing biological effects. Nevertheless, theoretical modeling still would claim that most residential and occupational epidemiological results are "impossible" (Weaver et al., 1998). It would also claim that bioeffects from magnetic field experiments using intensities less than 100 mG* are "impossible" (Adair, 1999). A milliGauss (mG) is a commonly used measure of magnetic field strength. An average living room would have a 0.7 mG field. The standard international unit is a microTesla (μT). One μT equals 10 mG. Both units appear in this document. Those who adhere to these biophysical theories still discount the relevance of experimental results at higher intensities because of this "impossibility" threshold and would require robust bioeffect laboratory results from ambient levels of exposure. This is an unusual burden of proof since ambient levels of other pollutants often do not produce effects large enough to see in the laboratory. It should be noted that the majority of panelists at IARC, NIEHS, and NRPB who declared EMFs as "possible" carcinogens obviously did not accept some physicists arguments that bioeffects from high-end residential exposures were "impossible."

Mechanistic Research: EMFs, particularly those above 1000 mG, have been shown to have a number of physiological effects on cells (Portier & Wolfe, 1998), but the physical induction mechanisms of these effects are not clearly understood. No consensus has arisen on a mechanistic explanation of how the various epidemiological associations might have occurred. Repeated studies of the effects of pulsed and non-pulsed EMFs below 100 mG on chick embryos, in several laboratories, have continued to show "non-robust" effects (Martin, 1988), (Berman et al., 1990), (Martin, 1992), (Moses & Martin, 1992), (Moses & Martin, 1993), (Martin & Moses, 1995), (Litovitz et al., 1994), (Farrell et al., 1997a), (Farrell et al., 1997b), (Leal et al., 1989), (Chacon et al., 1990), (Ubeda et al., 1994), (Koch & Koch, 1991), (Singh & et al., 1991), (Espinar et al., 1997), (Blackman et al., 1988), (Yip et al., 1994a), (Yip et al., 1994b), (Coulton & Barker, 1991), (Youbicier-Simo et al., 1997), (Piera et al., 1992), (Pafkova & Jerabek, 1994), (Pafkova, Tejnorova & Jerabek, 1994), (Pafkova et al., 1996), (Veicsteinas et al., 1996). A statistically significant effect is said to be "non-robust" when its size is not greater than the differences between control groups in various experiments. Several independent researchers (Liburdy et al., 1993), (Blackman, Benane & House, 2001), and (Ishido, Nitta & Kabuto, 2001) have published studies on the effect of low intensity (12 mG, 60 Hertz) magnetic fields on the ability of melatonin to inhibit cancer cell proliferation in vitro. Thus, there are some studies that, while not universally accepted, purport to show biological effects at EMF intensities declared by biophysicists to be incapable of producing such effects.

Animal Pathology: A large number of animal pathology studies have been carried out that tested a few aspects of the EMF mixture and, with some exceptions, did not show a carcinogenic, reproductive, or immunological effect (Portier & Wolfe, 1998). This has led some scientists to conclude that EMFs are probably safe.

Two laboratories in the former Soviet Union (Beniashvili, Bilanishvili & Menabde, 1991), (Anisimov et al., 1996) and one in Germany (Loscher et al., 1993), (Mevissen, Lerchl & Loscher, 1996a) reported co-promotional effects of magnetic fields on the occurrence of breast tumors in rats, though this result did not recur in two experiments in the United States (Anderson et al., 1999), (Boorman et al., 1999a) that partially replicated the conditions in the German experiments.

Epidemiology: Epidemiological studies on workers and children have tentatively implicated a wider range of diseases than the leukemia and brain cancer that dominated discussion in the early 1980s and 1990s (Portier & Wolfe, 1998). Published statistical summaries of the body of epidemiological evidence have suggested that chance is an unlikely explanation for the associations seen for childhood leukemia (Greenland et al., 2000), (Ahlbom et al., 2000), adult leukemia (Kheifets et al., 1997a), adult brain cancer (Kheifets, 2001), male breast cancer (Erren, 2001), and Amyotrophic Lateral Sclerosis (Ahlbom, 2001). This leaves bias, confounding, or EMF causality as alternative explanations. (See pp 21-22 below for definitions.) Parts of this evidence have convinced the NIEHS, the IARC, and the NRPB that EMFs are a **possible** carcinogen.

For childhood leukemia, the association now seems more consistent with measured 30-300 Hz magnetic fields than with proximity to power lines (Greenland et al.,

* A milligauss (mG) is a measure of magnetic field intensity. A typical living room measures about 0.7 mG. The average exposure during the day of a typical white-collar worker would be around 1 mG, a utility worker exposed to high fields during the day might average around 7 mG, while an electric train operator's exposure might average around 100 mG.

2000). Furthermore, alternative explanations of the associations, such as traffic and social class, seem much less likely (Reynolds et al., 2001), (Langholz, 2001). The study of Linet et al. on childhood leukemia (Linet et al., 1997) was originally and prominently interpreted as showing no effect. It has now been shown to contribute important support in pooled analyses that indicate that the association between the highest exposures to EMF and childhood leukemia are unlikely to be due to chance (Greenland et al., 2000).

An epidemiological literature is developing that associates magnetic fields with diseases and conditions that are more common than cancer, such as sudden cardiac death, dementia, suicide (NIEHS, Portier & Wolf, 1998), and spontaneous abortion (Li et al., 2002), (Lee et al., 2002). From a cost/benefit perspective, the confirmation of the associations with these more common diseases would have greater utilitarian policy implications (Florig, 2001) than the confirmation of EMF associations with rare diseases, such as childhood cancer or Lou Gehrig's Disease (amyotrophic lateral sclerosis).

Exposure: A number of epidemiological studies and exposure surveys have given a significantly better description of the range of exposures to some aspects of the EMF mixture, both in the occupational and in the general environment (Portier & Wolfe, 1998), (Li et al., 2002), (Lee et al., 2002), (Zaffanella & Kalton, 1998), (Zaffanella & Hooper, 2000). It has become clear that the 24-hour average of the minute-by-minute 50-60 Hz magnetic field exposures is primarily influenced by stray ground currents, internal wiring, and the power grid rather than by appliances. Maximum fields (the highest exposure during the day) are probably contributed by use of appliances, electrical transportation, or passing briefly by internal wires, current-bearing plumbing, or very close to above or below ground power lines.

Which Aspects of the "EMF Mixture" Might Be Bioactive?: As the decade of the 1990s began, a few childhood leukemia studies suggested that associations were stronger between leukemia and proximity to power lines than between the disease and measured fields (NAS et al., 1997). With more studies, this pattern has disappeared (Greenland et al., 2000). The earlier impression led to investigations of correlates with power lines and measured magnetic fields. Resonance between the static magnetic field of the earth and alternating 60 Hz fields was evaluated, as were transient changes in magnetic field, as potential explanations for the epidemiology. As indicated on page 32, the results do not strongly implicate these aspects of the EMF mixture (Kaune et al., 2002).

A new hypothesis has arisen (Kavet et al., 2000), (Dawson et al., 2001). It proposes that contact currents from low frequency voltages, and not exposure to magnetic fields, might explain some of the epidemiological associations. Others (Graham and Ludquist personal communication, 2001) suggest that the high frequency components of these currents are bioactive. In occupational settings, micro-shocks have been invoked to explain the persistent association between magnetic field exposure and ALS (NRPB, 2001b), (Ahlbom, 2001). These hypotheses have not yet been tested.

Scattered associations with electric fields have been reported (Coghill, Steward & Phillips, 1996), (Miller et al., 1996), but this association has not been consistent. A hypothesis and some evidence have developed with regard to electric fields near transmission lines and their effects on the charge and concentration of particulate air pollutants (Henshaw et al., 1996). If true, this would suggest that one should bury lines to block their electric fields and that rephasing would not be effective. However, this hypothesis has not been sufficiently supported by evidence.

Two recent studies of miscarriage and personal EMF exposure suggest that maximum fields or average change between consecutive exposures may convey risk (Li et al., 2002), (Lee et al., 2002). Studies of the effect of personal exposure on urinary melatonin metabolites in utility workers have suggested the possibility that the rate of change of the magnetic field may be bioactive (Burch et al., 1998). This, too, would have implications for any mitigation. One laboratory has reported that the super-imposition of random EMF noise in the laboratory can block the effects of orderly low-frequency magnetic fields (Litovitz et al., 1994). No replication of this study has been attempted yet.

Radio Frequency Research: Public concern and research on the question of radio frequency and low-frequency-modulated radio frequency have increased in the last decade. Although this area may turn out to be relevant to the low frequency literature reviewed here, exploration of it was beyond the resources, mandate, and expertise of the review team.

Funding: Funding for EMF research in the United States has dropped from the levels in the late 1980s. The Department of Energy research program of $10 million per year has been eliminated and the amount of resources devoted to EMF research by the utility industry and the Electric Power Research Institute has decreased from $10 million per year at its peak to $3.5 million in 2000. The National Institutes of Health have no special study section with EMF experts to review research proposals in this area, so proposals are judged by experts in other areas and compete for scarce research dollars.

3 How to Read This Document

This document is not just a summary of the facts from the vast literature on the possible health effects of extremely low frequency (ELF) electric and magnetic fields. Instead the bulk of the main document presents a much more detailed rationale for the conclusions drawn, and the evidence is summarized in graphical and tabular form.

In preparation for this evaluation, the California EMF Program held a two-day epidemiology workshop to discuss some of the most relevant epidemiological findings and methodological issues. The proceedings of that workshop, which were pivotal to some of the conclusions reported here, were published in a peer-reviewed Supplement (5) of the journal *Bioelectromagnetics* on January 22, 2001.

4 What is New in this Evaluation

New Evidence

There have been many adequate reviews, including some very recent ones (NAS et al., 1997), (Portier & Wolfe, 1998), (IARC, 2001). The NIEHS review, in particular, was regarded as the starting point for this evaluation. The NIEHS Working Group carried out their evaluation in June 1998. Several important studies have been published between the conclusion of the NIEHS Working Group review and this evaluation, including three major studies on childhood leukemia (Green, Miller & Agnew, 1999b), (Green et al., 1999a), (McBride et al., 1999), (UKCSS, 1999). The deadline for including studies in our evaluation was June 24, 2000. This is later than the deadline originally mentioned in the Risk Evaluation Guidelines (REGs). Since the DHS evaluation began later than initially envisaged, the reviewers felt that it was unwise to disregard recently published, and possibly important, studies simply to observe a previously set but otherwise arbitrary date. Only one large study (van Wijngaarden et al., 2000) that dealt with suicide emerged during this extended deadline period.

In addition, the reviewers considered studies sponsored by the California EMF Program (Li et al., 2002), (Lee et al., 2002) and in the Epidemiology Workshop satisfying the criteria for inclusion in this evaluation, as specified in the Guidelines. In this final draft, the DHS scientists also discuss articles that were brought to their attention during the public comment period.

The document has features that were not present in the NIEHS document. One of these—presenting a graded degree of certainty of causality—was described above. Also discussed are the aspects that make up the EMF mixture that characterizes the exposure of persons who come near the power grid, the internal wiring of houses, and common household appliances. These are described in Chapter 3. The reviewers stress the notion of "mixture" because different aspects of EMF exposure (e.g., 60-cycle magnetic fields and high-frequency transients) would require different actions for abatement. For each of the diseases considered, there are explicit discussions about whether the epidemiological associations observed, if real, would convey a risk from lifetime exposure that would be of regulatory interest. This is a parameter of interest to the social justice policy framework, which focuses on the individual risks of the most highly exposed. In Table IX, the baseline mortality for conditions considered possibly associated with EMFs are discussed. The reviewers ask if the attributable burden of mortality from even a very small fraction of that baseline would be of regulatory interest when compared to the mortality burden thought to be avoided by regulation of other agents. The attributable burdens of mortality or morbidity are parameters of interest to the utilitarian policy framework, which aims at the most good for the most people at the least cost. The document also attends to any evidence suggesting inequitable exposure or vulnerability to EMFs. This is relevant to the environmental justice policy framework, which is concerned with unfair distributions of risk.

Each health condition considered had at least two epidemiological studies in which there was a statistical association with some surrogate for EMF exposure. The list of conditions is similar to that discussed in the NIEHS document and includes

- Adult and childhood leukemia
- Adult and childhood brain cancer
- Male and female breast cancer
- EMF as a "broad spectrum" carcinogen for all cancers
- Miscarriage
- Other reproductive and developmental conditions
- Amyotrophic lateral sclerosis (Lou Gehrig's Disease)
- Alzheimer's disease
- Acute myocardial infarction

- Suicide
- Other adverse non-cancer health outcomes (depression, electrical sensitivity)

5 QUALITATIVE BAYES OR DEGREE OF CERTAINTY APPROACH TO EVALUATION

The DHS scientists found the usual process of describing the pattern of evidence in some detail and then expressing an opinion (without explaining the rationale for that opinion) to be insufficiently transparent. Accordingly, they supplement the usual IARC procedure with an additional form of presentation and an additional form of judging whether EMFs are a cause of disease. The following table shows the questions that were systematically addressed. For definitions of epidemiological terms in the table see pages 20-22 (Sections 12.1.1-12.1.3).

TABLE II. QUESTIONS RELEVANT TO DEVELOPING A DEGREE OF CERTAINTY ABOUT CAUSALITY

EXPLANATIONS OF A STATISTICAL ASSOCIATION OTHER THAN A CAUSAL ONE
Chance: How likely is it that the combined association from all the studies of EMF and disease is due to chance alone?
*Bias: How convinced are the reviewers that EMFs rather than a study flaw that can be **specified and demonstrated** caused this evidentiary pattern? If no specified and demonstrated bias explains it, how convinced are they that EMFs caused these associations rather than **unspecified** flaws?*
*Confounding: How convinced are the reviewers that these disease associations are due to EMFs rather than to another **specified and demonstrated** risk factor associated with EMF exposure? If not due to a specified risk factor, how convinced are they that they are due to EMFs rather than to **unspecified** risk factors?*
*Combined effect: How convinced are the reviewers that these disease associations are due to EMFs rather than to a combined effect of chance and specified or **unspecified** sources of bias and confounders?*
ATTRIBUTES SIMILAR TO HILL'S (HILL, 1965) THAT ARE SOMETIMES USED BY EPIDEMIOLOGISTS TO EVALUATE THE CREDIBILITY OF A HYPOTHESIS WHEN NO DIRECT EVIDENCE OF CONFOUNDING OR BIAS EXISTS
Strength of association: How likely is it that the meta-analytic association is strong enough to be causal rather than due to unspecified minor study flaws or confounders?
Consistency: Do most of the studies suggest some added risk from EMFs? How likely is it that the proportion of studies with risk ratios above or below 1.0 arose from chance alone?
Homogeneity: If a large proportion of the studies have risk ratios that are either above or below 1.0, is their magnitude similar (homogeneous) or is the size of the observed effect quite variable (heterogeneous)?
Dose response: How clear is it that disease risk increases steadily with dose? What would be expected under causality? Under chance, bias, or confounding?
Coherence/Visibility: How coherent is the story told by the pattern of associations within studies? If a surrogate measure shows an association, does a better measurement strengthen that association? Is the association stronger in groups where it is predicted? What would be expected under causality? Under chance, bias, or confounding? How convinced are the reviewers that the magnitude of epidemiological results is consistent with temporal or geographic trends?
Experimental evidence: How convincing are the experimental pathology studies supporting the epidemiological evidence? What would be expected under causality, bias, chance, or confounding?
Plausibility: How convincing is the mechanistic research on plausible biological mechanisms leading from exposure to this disease? What would be expected under causality, chance, bias, or confounding? How influential are other experimental studies (both in vivo and in vitro) that speak to the ability of EMFs to produce effects at low dose?
Analogy: How good an analogy can the reviewers find with similar agents that have been shown to lead to similar diseases? What would be expected under causality, chance, bias, or confounding?
Temporality: How convinced are the reviewers that EMF exposure precedes onset of disease and that disease status did not lead to a change in exposure?
Specificity and other disease associations: How predominantly are EMFs associated with one disease or subtypes of several diseases? What would the reviewers expect under causality, chance, bias, or confounding? How much is their confidence in EMF causality for disease X influenced by their confidence that EMFs cause disease Y?

As a heuristic device, and following Huticinson and Lane (Hutchinson & Lane, 1980), the REGs suggested that these questions about the pattern of evidence be posed so that one could say the pattern is more likely under the hypothesis that EMFs contributed to the cause of that health condition or more likely under the hypothesis that chance, bias, or confounding produced the pattern. This allows the reviewers to provide the reader a rationale for the relative weight given mechanistic, animal pathology, and epidemiological evidence and to understand which parts of the evidence suggest causality and which speak against causality.

The DHS reviewers coined the term "Qualitative Bayes Approach" to characterize a form of verbally justifying judgments about hazard that paid attention to the insights of Thomas Bayes, an 18th-century mathematician. His insights would suggest starting with some initial degree of certainty that any given agent is capable of being harmful based on knowledge about agents in general. Evidence is then accumulated on this specific agent and this changes the degree of suspicion or certainty. Imagine a prehistoric hunter deciding whether to try out some jungle fruit he has never seen before. He has an initial degree of suspicion high enough that he does not partake right away. He takes some fruit home and feeds it successively to several types of captured birds. As each species seems to survive, it seems less and less likely that the fruit would be harmful to humans. But since the leaves of the tree bearing that fruit resemble those from a tree that bears a poisonous fruit (causing the initial suspicion to be very high) the hunter's specific experiments might still leave him fairly suspicious and lead him to cruelly feed the fruit to a captive from another tribe. Only if the captive survived would his initial suspicions be allayed. This example illustrates Thomas Bayes's two key insights. As evidence builds we update our degree of certainty of harm, but, at any point in time, that updated degree of certainty also depends on how suspicious we were initially. This idea is expressed mathematically by a simple formula. The first term of the Bayes formula is the "prior odds," that is, the odds that a given hypothesis is thought to merit *a priori*, before examining the evidence. In this document it is called the prior because it is not based on subsequent research.

The second term, the "likelihood ratio," is a multiplier, calculated (or, in this case, qualitatively discussed) after scientific evidence has been collected and evaluated. The term "likelihood ratio" is most properly restricted to the case where one compares the statistical likelihood of a result under one specific hypothesis relative to that under another hypothesis, usually the null. It expresses the likelihood of the observed pattern of evidence if EMFs do indeed cause disease, divided by the likelihood of that pattern if EMFs do not cause disease. The third term, the "posterior," is the product of the first two and represents the odds of the risk being true after the prior has been modified by our evaluation of the evidence.

Because of the difficulty of translating complex evidence into numbers, we only use the ideas behind the formula as a way of explaining how certain or uncertain we were to begin with and to explain the basis for the weights we gave a particular stream of evidence in order to update our degree of certainty. The Bayesian perspective used by the California reviewers recognizes that a reassuring pattern of evidence from a stream of evidence that often misses a harmful effect does not allay one's suspicion much, even though an alarming pattern of evidence from that same stream of evidence might increase suspicion a lot. Going back to the hunter-gatherer example: if birds sometimes survive eating fruits that are lethal to humans, then reassuring evidence from bird experiments would not allay suspicion as much as the death of the birds after eating the fruit would increase our suspicion. In the terminology of probability, the relative likelihood conveyed by a positive or negative result depends on the false-positive rate and false-negative rate characteristic of that stream of evidence. The mathematical basis for this insight is discussed in the REGs (www.dhs.ca.gov/ehib/emf). It resulted in realizing that any stream of evidence, judged by the extent to which it usually produced false-positive and/or false-negative results, could be classified into four possible types: 1) capable of strengthening OR weakening one's certainty, 2) predominantly capable of strengthening certainty (like the bird feeding example given above), 3) predominantly capable of weakening certainty and, 4) uninformative, neither capable of strengthening nor weakening one's confidence. While this structured discussion helped organize the reviewers' judgments, it did not involve a mathematical combination of weights as would be the case in a quantitative Bayes evaluation. It should be noted that the Hill's attributes are like the bird-feeding example. If they are present they strengthen confidence, but if they are absent, confidence falls only a little.

The DHS reviewers considered the following streams of evidence: biophysical evidence about the physical induction mechanism, research into physiological and pathophysiological mechanisms, research into animal pathology and epidemiological evidence. Clearly if all these streams of evidence were non-supportive, one's degree of certainty would fall, and if they were all supportive it would rise. If some streams of evidence are unsupportive and some are supportive, the DHS reviewers considered the inherent proclivity of each stream of evidence to give false positive or false negative results as a guide to what weight its results should be accorded. If apparently supportive evidence is shown clearly to be due to artifacts, this would lower the degree of certainty.

In the "Qualitative Bayes Approach" the DHS reviewers elicited their own expert judgment about the *a priori* (initial) probability of hazard after a special training session on how to avoid common errors of probabilistic estimation. It was important

to be explicit about the prior probability because some physicists were arguing on the basis of physical theory applied to simplified biological models of the cell, that any biological effect from residential EMFs was impossible and thus had a vanishingly small initial credibility. This meant that they would require extraordinarily strong specific evidence to change their initial impression. Previous risk assessments have not explicitly considered this issue.

The discussion then turns to the patterns of specific EMF evidence in biophysical, mechanistic, animal pathology, and epidemiological streams of evidence. Obviously, if all four streams of evidence pointed toward or away from an EMF effect, the reviewers' job would be easy. But what if some streams of evidence are supportive and some are not supportive? What weight should be given each stream of evidence? It was in the effort to address this problem that discussions of the inherent proclivity to give false positive and negative results came into play. This discussion was guided by a series of pre-agreed-upon questions described in the table above. The discussion included pro, con, and summary arguments. An example of such arguments are presented in the next table.

TABLE III. EXAMPLE OF PRO, CON, AND SUMMARY ARGUMENT

CHANCE		
AGAINST CAUSALITY	**FOR CAUSALITY**	**COMMENT AND SUMMARY**
(A1) Not all the associations (relative risks) are above 1.00 or statistically significant.	(F1) The narrow confidence limits in the meta-analytic summaries and the low likelihood of this pattern of evidence by chance leans away from chance as an explanation.	(C1) A non-chance explanation must be sought.

Considering this kind of structured discussion helped organize the reviewers' judgments, after he/she weighed all the information in the usual way, although it did not involve a mathematical combination of weights as would be the case in a quantitative Bayes evaluation. After consideration of this carefully structured discussion of the evidence (considering how much more—or less—likely the pattern of evidence would be if the risk hypothesis were true compared to the likelihood of that evidence if EMFs were safe), the reviewers expressed an expert judgment on the posterior probability of a causal relationship.

6 QUALITATIVE BAYES RISK EVALUATION COMPARED TO TRADITIONAL AND QUANTITATIVE BAYES RISK EVALUATIONS

The traditional risk assessment has a section in which a judgment is given as to whether the agent being evaluated is capable of causing cancer or some other adverse health effect. This is called the "hazard identification." The typical presentation is heavy in describing the relevant evidence and rather light in explaining the rationale for the conclusion. Often the weight, given mechanistic, animal pathology, and epidemiological streams of evidence, depends on a review panel's interpretation of adjectives which best describe the pattern of evidence. For example, is the pattern of evidence "sufficient" or should it be called "limited"? Can confounding and bias be "reasonably" discounted? Then there are pre-agreed-upon rules for combining the streams of evidence. Limited animal evidence plus limited epidemiological evidence results in one rank, sufficient animal evidence plus limited epidemiological evidence leads to another rank, and so forth. The combinatorial rules are straightforward, but the rationale for deciding that a stream of evidence is "limited" is not clearly defined and is subjective.

A completely quantitative Bayesian approach of the sort proposed by McColl et al. (McColl et al., 1996) or by Lindley (Lindley, 2000), would require assigning many quantitative parameters to a complex Bayesian Net model which would mathematically combine the subjectively assigned parameters to produce a posterior degree of certainty of causality. To the reviewers' knowledge, this kind of model has never been applied to any environmental agent. How experts such as physicians, combine streams of evidence to make judgements about causality has

been of great practical interest. As pointed out by Shortliffe (Shortliffe et al., 2001) there have been two general approaches. One is to infer statistically (Holman, Arnold-Reed & Klerk, 2001) or find by interview what rules experts usually employ. This assumes that the rules of thumb that experts use are optimal. As Holman (Holman et al., 2001) points out, however, this may not always be the case. The other approach is to use information to indicate what weights ought to be used. An example of this was de Dombal's (de Dombal et al., 1972) work using a Bayesian approach to diagnosing the acute abdomen on the basis of the prior probability of patients with certain diagnoses showing up in emergency rooms, and the relative likelihood of elements of medical history, physical signs, and laboratory test results in the several possible diagnoses. According to Shortliffe (Shortliffe et al., 2001), neither approach has so far been reduced to computer applications that render the combining of streams of evidence a cut and dried uncontroversial activity. It should be expected then, that the analogous task of risk evaluation will still rely on professional judgement and will not be free of controversy. For this reason, our stakeholders urged us to opt for transparency rather than computational elegance in our risk evaluation guidelines. In response to the third draft, the Electric Power Research Institute contracted with Professor Sander Greenland in late 2001 to prepare a quantitative Bayesian model based on the epidemiological evidence for childhood leukemia. Since his will be the only extant quantitative Bayesian epidemiological analysis, the reviewers contrast its proposed approach to their own. His model will provide a posterior dose-response curve based on a prior dose-response curve, the pooled epidemiological data, and prior estimates of selection bias and non-differential measurement bias. The all-important biophysical, mechanistic, and animal pathology streams of evidence will not be part of Greenland's model, although they could influence the prior dose-response curve in a subjective way. Calculations from Greenland's model would allow one to provide a probability that the posterior slope of the dose-response curve is not flat, that is, that there is some causal effect.

The following table compares the Qualitative Bayes evaluation to the traditional and to Greenland's Quantitative Bayes approach to risk evaluation as to a number of characteristics.

TABLE IV. COMPARISON OF USUAL RISK ASSESSMENT METHOD TO QUALITATIVE AND QUANTITATIVE BAYES METHODS

CHARACTERISTIC	USUAL METHOD	QUAL. BAYES	QUANT. BAYES
Evaluates all streams of evidence?	Sometimes	Yes	Focuses on epidemiology, other streams influence prior
Elicits prior probability?	No	Yes	Prior dose-response curve
Compares likelihood of each element of the evidence under the hazard and non-hazard hypotheses?	No	Qualitatively	Quantitatively with many of the parameters subjectively elicited
Pro, con, and summary arguments to make rationale transparent?	No, most risk assessments are skimpy in justifying hazard categories assigned	Yes	Not unless a supplementary document were to accompany the model
Combines relative likelihoods mathematically to derive posterior?	No	No	Yes, but in some versions non-epidemiol. evidence is folded into the prior subjectively
Elicits an expert posterior probability after considering all	No	Yes	No

CHARACTERISTIC	USUAL METHOD	QUAL. BAYES	QUANT. BAYES
elements of the evidence?			
Displays judgments of various judges separately?	Usually strives for semblance of consensus	Yes	Technically possible for different experts to elicit their own parameters
Frames intermediate degrees of certainty as "not a proven hazard?"	Often	No, reveals posterior probability	No, reveals posterior probability

Both the Qualitative Bayes and the Quantitative Bayes evaluations can provide a posterior degree of certainty that the epidemiological associations are causal, which, if in the range from 10 to 90 out of 100, will not seem trivial to the general public and will stimulate policy discussions. The statements, "possible," "there is no proven hazard," or "there is no consistent evidence," often used for this range of degrees of confidence, will not stimulate such discussions. Thus, both the Qualitative Bayes and Quantitative Bayes methods pose risk communication "problems" for those who believe that society should not begin policy discussions until most scientists are virtually certain that a hazard exists. The traditional hazard identifications would pose the same "problem" if they routinely used more nuanced categories of hazard assessment that distinguished between, say, a certainty level of 11/100 and one of 89/100. As now framed they pose a risk communication "problem" for those who believe that policy discussions should begin even before a hazard is firmly established.

Compared to traditional qualitative evaluations, the Qualitative Bayesian approach makes the evaluation more transparent, but it still accommodates different opinions. The DHS reviewers have no doubt that critics of their conclusions could use the Qualitative Bayes format to make their points. Some of the physicists who believe that they have a theory to prove that no residential EMF effect is possible would use priors so low that their posterior degrees of certainty would be low as well; the toxicologists who believe reassuring animal tests prove that EMFs are safe would make a case that the animal study results pull down their degree of certainty of a hazard to a level below their initial degree of certainty. In a contentious area such as EMFs, the reviewers doubt very much that any of the three styles of risk evaluation discussed in the table would force a consensus among subject matter experts who weigh and interpret the several streams of evidence differently. Even in the Quantitative Bayes model experts will use different priors and will elicit different subjective relative likelihood parameters for items like bias and confounding, for which there is no direct evidence. In the traditional method, experts will disagree on whether a stream of evidence warrants the adjective "limited" or "sufficient," and in the Qualitative Bayes approach experts will disagree on "how much more likely" the pattern of evidence is under the causal and non-causal hypotheses. But the reasons for these different judgments will be more transparent in the Qualitative Bayes style of risk evaluation and we believe that this is desirable in controversial areas.

7 How Credible Was the EMF Hypothesis to Begin With?

The three reviewers first considered the initial credibility of the hypothesis (before any targeted research had been done) that everyday residential and electrical occupational EMF exposures could influence the risk of disease. Like the majority of reviewers at IARC and NIEHS, the DHS reviewers were swayed only a little by theoretical biophysical arguments that such influences were impossible, since these arguments depend on assumptions about biological systems that may or may not be sophisticated enough to reflect reality and rule out an effect. The reviewers acknowledged, though, that this was probably the only agent they had encountered where these kinds of "impossibility" arguments had been made. However, a better understanding of biology (and not any change in physics theory) could conceivably explain how an organism could detect and be affected by the spatially and temporally coherent EMFs or other aspects of the EMF mixture emanating from power lines and appliances.

The reviewers considered the proportion of chemical agents that had tested positively for carcinogenicity at high doses (about 20%) as one benchmark (Fung et al., 1993). They also considered the fluctuation of disease rates starting in the late 19th century when electricity began to spread gradually from wealthy urban areas to other parts of the world. Any changes could put *a priori* bounds on the size and direction of any EMF effect. Milham (Milham & Ossiander, 2001) drew attention to

something that Court Brown and Doll (Brown & Doll, 1961) had pointed out more than 40 years ago, that an increased risk of leukemia mortality for 2- to 4-year-old children first appeared in the 1920s and increased in intensity in the 1940s. Thus some factor(s) (perhaps electricity, perhaps accuracy in diagnosis), in those modernized locations caused the registration of toddler leukemia deaths to increase threefold. The evidence from Court Brown, Doll, and others that childhood leukemia mortality registration had indeed increased during the early 20th century increased the prior probability of a moderately large EMF effect, at least for childhood leukemia. Since similar trends were not reported for other conditions, it was considered that modest protective or harmful effects from rare high exposures were compatible with the data.

The three DHS reviewers underwent special training in probability elicitation. They then judged that EMF effects were about as probable or a little less probable to influence the risk of disease as any man-made environmental pollutant taken at random. The three reviewers gave probabilities ranging from 5% to 12% *a priori*, that EMFs at or above the 95th percentile of typical residential US exposures would produce effects detectable by epidemiologists when compared to the 1st percentile of residential exposure or below.

8 The Weight Accorded Biophysical Arguments that Bioeffects from Residential and Most Occupational Fields were Impossible or that No Physical Induction Mechanism had Been Elucidated

While the reviewers do not doubt established physical theory, they believe that its application to simplified biological models is not sufficiently convincing to prove the impossibility of epidemiological or laboratory observations. However, the argument that environmental fields have very little energy lowered the prior probability that EMFs might have biological or pathological effects. The fact that there was no mechanistic explanation for how residential-level electric or magnetic fields might cause chemical or cellular changes, that there was no recognized molecule or organ capable of reacting or detecting residential magnetic fields, and the fact that recognized physiological effects of pulsed and very high magnetic fields did not have a well-understood physical induction mechanism did not decrease the updated degree of confidence much. This is because many known physiological and pathological effects go for a long time without a full mechanistic understanding.

9 The Weight Accorded Experimental Evidence On Any Pathophysiological Mechanisms by Which EMF Might Work

It has long been known that EMFs can affect biological processes, if their intensity is strong enough. In fact, safe exposure limits have been set to prevent these effects. A good review can be found in the book *Electromagnetic Fields (300 Hz to 300 GHz), Environmental Health Criteria 137*, published under the joint sponsorship of the United Nations Environment Program, the International Radiation Protection Association, and the World Health Organization (Geneva, 1993). In almost all cases, these levels are exceeded only in very rare occupational environments. Since they are almost never exceeded in the general environment, such levels are not a public health concern. A much more complex debate centers on whether these are the only possible effects or whether the temporal and spatial coherence of the man-made fields associated with electric power can be somehow discriminated from the incoherent endogenous currents and interact with biological processes at levels much lower than those for which exposure limits exist. The reviewers agreed that, as was also the case initially for many disease-causing agents, there is not a well-documented mechanism that explains how the EMF "mixture" at residential or occupational levels could initiate a biological response or, having initiated that response, how a chain of events could lead to damage or disease of various types. There are biological effects from aspects of the EMF mixture, particularly at exposure doses far above residential and occupational levels. At this time they do not provide a clear mechanistic understanding of how the EMF mixture could cause disease. The absence of a clear mechanistic chain of effects and the failure of many experiments with aspects of the EMF mixture to produce any mechanistic effects did not lower the reviewers certainty of causality much below what it was initially. The evidence that there are some mechanistic effects of some aspects of the EMF mixture at doses (thousands of mG) far higher than usually encountered in the environment did not boost the confidence of causality very much beyond the initial probability because the biophysical arguments suggest that they might not be relevant to effects at lower levels. The DHS reviewers accepted the unusually strict requirement that mechanistic results in the laboratory must be demonstrable at ambient levels of exposure.

It should be noted that the assumption of many of the mechanistic experiments is that the effects of magnetic or electric fields (like those of many chemicals and ionizing radiation) occur at a level of organization demonstrable in a chemical mixture, a mixture of cellular components, or a mixture of cells and does not depend on the presence of an intact multicellular organism. There are some well-recognized effects that violate these assumptions. For example, the intact shark, through a

special organ with an array of connected detectors, can detect tiny electrical fields emitted by distant prey. The exact biophysical mechanisms by which the individual detectors work cannot be documented using individual receptors at the ambient levels detected by the intact shark (Kalmijn, 1971), (Wissing, Braun & Schafer, 1988).

The lack of mechanistic understanding, which was initially the case for many harmful agents, is not as strong an argument against causality as the presence of such an understanding would be in favor of causality. Therefore the mechanistic line of evidence did not contribute much to the reviewers' judgments.

10 The Weight Accorded to Experimental Evidence not Clearly Connected With Particular Endpoints but Relevant to the Ability of Low-Level EMFs to Be Bioactive

A number of studies, both in vivo and in vitro, report bioeffects which, while they do not shed light on physical induction or pathophysiological mechanisms, do suggest that there are effects other than those mediated by well-understood mechanisms, such as induced currents. For example, the initial observations by Liburdy of inhibition of the melatonin antiproliferative action by 12 mG 60 Hz fields in 1993 (Liburdy et al., 1993) has been confirmed and extended by two other laboratories (Blackman et al., 2001), (Ishido et al., 2001). The series of studies using pulsed magnetic fields that showed non-robust effects on chicken embryos at intensities below 100 mG (Martin, 1988), (Berman et al., 1990), (Martin, 1992), (Moses & Martin, 1992), (Moses & Martin, 1993), (Martin & Moses, 1995), (Litovitz et al., 1994), (Farrell et al., 1997a), (Farrell et al., 1997b), (Leal et al., 1989), (Chacon et al., 1990), (Ubeda et al., 1994), (Koch & Koch, 1991), (Koch et al., 1993), (Singh & et al., 1991), (Espinar et al., 1997), (Blackman et al., 1988), (Yip et al., 1994a), (Yip et al., 1994b), (Coulton & Barker, 1991), (Youbicier-Simo et al., 1997), (Piera et al., 1992), (Pafkova & Jerabek, 1994), (Pafkova et al., 1996), (Pafkova et al., 1994), (Veicsteinas et al., 1996) also provide some evidence of bioeffects that would be considered “impossible” according to biophysical theory. These two areas of research have been greeted with suspicion. For example, Weaver (Weaver, Vaughan & Martin, 1999) dismisses in vitro effects as being artifactual, due to an insufficiently rigorous lack of temperature control, because biophysical theory suggests that tiny fluctuations in temperature would produce more effects than magnetic fields below 100 mG. The DHS reviewers were not convinced by this argument. These studies were no less rigorously conducted than most in vitro studies in other fields of research. There is no direct evidence that inducing magnetic fields also heats the tissues. If experimental controls beyond the current technological limits are required, then ALL in vitro and in vivo research should be called into question.

The reviewers had differing opinions on the extent to which this evidence should change the belief in the hypothesis from what it was when this issue was first raised. One could argue that any experiment that shows an effect where none is expected ought to increase the credibility that EMF can indeed interact with biological systems at energy levels that biophysical theory considers too low to be effective. These studies thus provide some grounds for mistrusting the prediction of simplified biophysical models that no effect is possible below 100 microTesla (μT). Reviewer 1 was compelled by the evidence as it stands, while the other two reviewers would require further experimentation to gain general acceptance of the results before putting a lot of weight on them. All three reviewers agreed that confirming or explaining away the results from these two groups of experiments would be important for those who put great weight on biophysical “impossibility” arguments.

11 The Weight Accorded to Animal Pathology Experiments

The reviewers agreed that, with few exceptions, animal pathology studies based on high exposures to certain aspects of the EMF mixture showed no effects. There were three reasons why the reviewers believed that animal bioassays of single ingredients of the EMF mixture might be prone to missing a true effect:

a) Finding the right animal species to test: While the reviewers recognized that most agents found to cause cancer in humans also cause cancer in some (but not all) animal species, they were also cognizant that there are known human carcinogens, such as cigarette smoke, alcoholic beverages, benzene, and arsenic, for which no animal model existed for many decades.

b) Testing one ingredient of a mixture: The reviewers all questioned whether the bioassay of one element of a mixture could be sensitive enough to detect problems in the entire mixture. For example, many reassuring assays on the carcinogenicity of caffeine would not reassure us about the carcinogenicity of coffee. The animal pathology studies to date have been on pure steady 60 Hz fields not on the mixture of ingredients found near power lines or appliances.

c) Assuming that high intensities of magnetic fields produce larger effects than moderate fields do: The reviewers also questioned the sensitivity of a bioassay involving a small number of animals and assuming a monotonically increasing risk from low to high-dose, when the epidemiological studies that prompted the bioassays did not suggest an ever-increasing response.

The epidemiology suggests that the effect, if any, at 100s of mG (Tynes, Reitan & Andersen, 1994b), (Floderus, Tornqvist & Stenlund, 1994), (Alfredsson, Hammar & Karlehagen, 1996), (Minder & Pfluger, 2001) is no greater than that of children at 3 mG (Greenland et al., 2000), or of highly exposed utility workers with 24 hr time weighted averages (TWAs) around 7 mG (Kheifets, London & Peters, 1997b), (Kheifets, 2001). One would not expect rodents at 1000 mG to demonstrate a large enough effect to be detected in a conventionally sized laboratory experiment with a few hundred animals.

Accordingly, the lack of response in most animal pathology studies did not lower the degree of certainty by much. Reviewer 1 and 3 had their degree of confidence increased somewhat by repeated, but unreplicated, results from one German laboratory (Mevissen et al., 1996b) and isolated results from two laboratories in the former Soviet Republics (Anisimov et al., 1996), (Beniashvili et al., 1991), which showed co-promotional effects on breast tumors. None of the reviewers were much influenced by the statistically significant increase in thyroid cancers in one of the bioassays (Boorman, McCormick & Findlay, 1999b), even though it had not appeared in control series of previous bioassays and was thus a very unlikely occurrence. This effect showed up in only one sex of rats and not in mice and thus did not pass conventional toxicological criteria for animal carcinogenicity.

12 The Weight Accorded to Epidemiology Combined with Other Streams of Evidence

In the reviewers' judgement, it was epidemiological evidence that produced the most change in the degree of certainty from what it was *a priori*. Epidemiological studies are non-experimental statistical studies of human populations that compare rates of disease in groups with different levels of exposure or compare the proportion of exposed subjects in groups of healthy and diseased persons. The weakness of epidemiological evidence is that one cannot rule out the effect of factors associated with EMFs ("confounders") or completely avoid the limitations of collecting evidence in the real world instead of a controlled laboratory environment. These limitations may introduce errors ("bias") in the results. On the other hand, the strength of epidemiology is that it deals with the species of interest (humans) and the mixture and dose of interest (the EMF mixture as experienced by humans).

The individual studies, most of which were described in the NIEHS report, have been summarized in tables and graphs in this report. A structured evaluation of the epidemiological evidence was carried out for each of the 13 endpoints and summarized with the classification used by IARC and also by a statement of the degree of certainty that the observed epidemiological associations were causal in nature. In evaluating the credibility of epidemiological evidence, it is common to consider whether the risk being studied is "biologically plausible" and if "experimental evidence" exists to support the epidemiology. The three reviewers followed this practice considering the impact on the epidemiological findings of mechanistic evidence and evidence about bioactivity at near ambient levels under the heading of "plausibility" and of the animal pathology under the heading of "experimental evidence." However, these non-epidemiological studies were discussed in detail in separate chapters.

12.1 Issues Relevant to the Evaluation of the Epidemiological Evidence

Epidemiological results, because of the limitations of the data collected in a "real world" environment, need to be evaluated with particular care. The three major concerns are the effects of chance, bias, and confounding.

12.1.1 Chance

Epidemiological studies are expensive. Moreover, in the case of EMF and cancer, it may be virtually impossible to find sufficient subjects with both a rare disease and the rare high exposures. The very well-conducted studies carried out in some Scandinavian countries are based on so few subjects that a single additional case of cancer would change their findings. It is possible to reduce the effect of chance findings by combining results from a number of studies in a meta-analysis or even to merge the data collected for different studies in one large data set (pooled analysis). For health endpoints such as childhood leukemia (Greenland et al., 2000), adult leukemia (Kheifets et al., 1997a), adult brain cancer (Kheifets, 2001), amyotrophic lateral sclerosis (Ahlbom, 2001), male breast cancer (Erren, 2001), and miscarriage (Lee et al., 2002), (Li et al., 2002), pooled or meta-analytic analyses achieve conventional "statistical significance." This could be interpreted as follows: If these were randomized experiments without the possibility of bias or confounding, the statistical associations found would not be expected to occur by chance in 5 or fewer experiments out of 100 replications, if there really was no effect. Of course, epidemiological studies are not experiments, and it would be unethical and impractical to experimentally subject large numbers of humans to potentially harmful agents. This leads to the consideration of bias and confounding.

12.1.2 Bias

Any source of error in collecting the data may introduce a bias, which is a reason why the apparent result might not be the truth. A very common bias results from errors in assessing the true exposure of the subjects to the agent of interest, in this

case EMFs. Provided exposure of cancer cases and healthy controls is not assessed differently, this bias on average results in an underestimate of the risk, if one exists. When comparing the health risk of subjects exposed above one value to that of subjects below that value, non-differential misclassification of exposure* would not, on average, show an association if one does not truly exist. However, it may inflate the risk of intermediate exposure subjects and thus frustrate attempts to estimate a dose-response function. In most of the EMF studies, measurements were not taken for a long enough duration during the induction period of the disease to avoid this kind of misclassification. And there is even some argument about whether the right aspect of the EMF mixture has been measured. The three reviewers concluded that all of this may have led to an underestimate of any true effect of high versus low exposures and may have frustrated the ability to develop an appropriate dose-response curve.

Of the many errors that can creep into epidemiological studies, one in particular has been a source of argument with regard to a subset of the EMF epidemiological studies. We are referring to "selection bias" in some of the case control studies. A case control study is analyzed by comparing a series of cases with a disease to a series of healthy subjects as to their EMF exposure. If the cases display a higher proportion of high EMF exposure than the controls, this suggests a causal effect of EMFs. If, however, the probability of being selected for study is influenced both by whether one has the disease AND whether one had a high EMF exposure, then an apparent difference will appear between the cases and the healthy controls, which is the result of this biased selection and the result does not reflect any true effect of EMFs on the disease. One way to recruit healthy subjects is random telephone contact. This method excludes subjects of lower socio-economic status (SES), who may not have a telephone. Experience has shown that healthy controls of lower SES are sometimes less likely to participate in epidemiological studies than upper class subjects. In some studies, lower class subjects are more likely to live in neighborhoods with nearby power lines (Bracken et al., 1998). Since cancer patients of all social classes are easier to recruit (through a cancer registry) and more likely to be interested in participating, the effects of non-representative control selection may distort the comparisons between cases and controls and, therefore, the study results. In the case of EMF, it is claimed that the fact that there are more subjects living close to power lines among the cancer patients than among the healthy controls could be due to the fact that low SES subjects are more likely to live close to power lines and they are underrepresented in the control group. This issue of possible selection bias in case control studies is a particular issue for the North American case control studies on childhood leukemia. Hatch (Hatch et al., 2000) indicate that the association between childhood acute lymphoblastic leukemia (ALL) and front door magnetic fields greater than 3 mG was 1.9 (1.1-3.27) among full participants in their study but fell to 1.6 (0.98-2.61) when 147 partial participants were included. Although this difference was well within sampling variability, she suggested that it might be evidence of the presence of a selection bias which might be even more extreme if non-participants had their front doors measured and had been included in the analysis. Hatch (Hatch et al., 2000) concluded that "while confounding alone is unlikely to be an important source of bias....selection bias may be more of a concern...in case-control studies." The Scandinavian studies relied on cancer registries and lists of citizens and did not require permission of the subjects so that selection bias was not a problem. Ahlbom (2001) has shown that the results of the two groups of studies are not much different. The pooled analysis of all the studies he dealt with showed a relative risk for exposures above 4 mG as 2.0 (1.3-3.1), while the results after excluding the US studies was 1.7 (1.0-2.8). That is, the confidence interval of the two risk estimates overlap, indicating that there may or may not be some overestimate of the effect of living near power lines in the American studies, but that even if these are excluded, the association remains statistically significant. In the pooled analysis by Greenland et al. (2001), there was an effect of power line proximity ("wire code"), as well as an effect of measured magnetic fields. This might indicate some selection bias for power line proximity. Nonetheless, magnetic fields come only partially from power lines. Internal wiring and currents on plumbing form an important source (Zaffanella & Kalton, 1998). The only evidence we know of that examines personal EMF exposure from all sources and its relation to social class (Lee GM & Li D-K, personal communication) does not suggest differences in personal EMF exposure in different social classes. The evidence linking EMFs and adult leukemia, adult brain cancer, Lou Gehrig's disease, and Li's prospective miscarriage study come largely from study designs where selection bias is not possible (studies where rosters of healthy workers or subjects of high and low exposure are followed until death or health outcomes are determined from available records without requiring subject cooperation). Thus, although selection bias may have distorted the associations between EMF and childhood leukemia in some of the studies, the three reviewers did not believe that it totally explained the childhood leukemia findings and selection bias was not even an issue in the bulk of the studies related to adult leukemia, adult brain cancer, ALS, or in one of the two recent studies on EMF and miscarriage.

* "non-differential misclassification of exposure" is said to occur when errors of measurement occur equally in cases of disease and in healthy controls.

12.1.3 CONFOUNDING

The term "confounding" is derived from the Latin "confundere," to melt together. Epidemiologists use the term when the impact of two risk factors "melt together" and must be disentangled. If heavy alcohol consumption and smoking are both known to cause esophageal cancer, and people who drink also tend to smoke, then the effect of drinking will confound the effect of smoking and vice versa. Therefore one must correct for this confounding in the way the data are analyzed. Sometimes the non-effect of a factor which conveys no risk at all is confounded with the true effect of another factor. For example, it has been suggested that people who live near power lines also live on busy streets with lots of traffic and air pollution. This argument suggests that the effect of air pollution on childhood leukemia was confounded with the non-effect of the power lines, and the power lines were falsely implicated instead of the air pollution. Two conditions must pertain for an agent to be a strong confounder of the EMF effect on the various diseases discussed in this report. That agent must be strongly correlated with EMF exposure and it must have an effect on the studied disease that is even stronger than the apparent effect of EMF. If it is weakly correlated with EMF exposure it must have an effect on disease that is very strong indeed if it is to make EMF falsely appear to have an effect. Langholz (Langholz, 2001) has examined the candidate confounders for childhood leukemia and their association with power line proximity wire code. He concluded that while something connected with the age of home was a possibility, factors like traffic density, ethnicity, and smoking were not likely confounders. Indeed, not all studies of traffic and childhood leukemia suggest it as a risk factor (Reynolds et al., 2001), but a recent study of traffic and power line proximity and childhood leukemia (Pearson, Wachtel & Ebi, 2000) did suggest that there might be a joint effect. Hatch (Hatch et al., 2000) examined a variety of socioeconomic, and other confounders, and concluded that together, or alone, measured confounders would distort the association with ALL by less than 15%. Hatch also found no association between residential mobility, magnetic fields, or leukemia unlike Jones (Jones et al., 1993).

Electric shocks have been invoked to explain the relation between high-exposure jobs in the utility industry and ALS (Ahlbom, 2001), (NRPB, 2001a). If this were confirmed, they might also be invoked to explain the adult leukemia and brain cancer associations on the as yet unproven assumption that shocks could somehow cause cancer. However, the literature linking shock to ALS, unlike much of the literature linking high-EMF exposure jobs to ALS, depends on subjects remembering shocks. They are thus more vulnerable to recall bias than the EMF studies. Some of the studies suggest a protective, not a harmful, effect (Cruz et al., 1999); (Kondo & Tsubaki, 1981), (Gunnarson et al., 1992) and the size of the harmful effects of shock are less than the high EMF job effect (Deapen & Henderson, 1986), (Savettieri et al., 1991). No published study has demonstrated a correlation between shocks and high-EMF exposure jobs. Studies are underway to see if grounding currents are associated with measured magnetic fields and power line proximity. The three reviewers felt that the evidence for the confounders that had been proposed for EMF exposure did not have strong support and therefore their degree of confidence was not decreased by the pattern of evidence.

12.1.4 COMBINED EFFECT OF CHANCE, BIAS, AND CONFOUNDING

Although each of these possibilities by itself is unlikely to explain the association between EMF and cancer, is it possible that a combination of the three may be responsible for an artifactual finding? The DHS reviewers considered this possibility and concluded that this is not a credible explanation when many studies of different design have reported similar results. It is not impossible that individual studies may have their result completely explained by an extraordinary coincidence in which independent unlikely events occur simultaneously. However, for many diseases considered here the general pattern of results is not critically dependent on accepting each individual study as reliable. For example, in the case of childhood leukemia, it has been repeatedly shown that, even if a few studies are excluded, the results of meta-analyses, pooled analyses, or sign tests are not significantly altered.

In conclusion, the DHS reviewers, to different degrees, concluded that chance, bias, and confounding are not probable explanations for the reported associations when they have been reported repeatedly by independent investigators. In addition, the DHS reviewers considered other criteria, notably the Hill's criteria for causality, keeping in mind that these are not to be considered as strict rules to follow. Apart from consistency, which, as noted above made them doubt the non-causal explanation for a few endpoints, none of the Hill's attributes, when applied to the pattern of evidence, influenced their degree of certainty by much.

The DHS reviewers recognize the size of the associations between EMF exposure and the various diseases studied are not so far above the resolution power of the studies that confounding and bias could be definitively ruled out as explanations. They recognized that there was rarely an orderly progression of increased risk within studies and that the effects reported for groups with dramatically high exposures like electric train operators did not display dramatically high risks when compared to those with low or moderate exposures. There are also examples where the statistical results are not completely coherent. However, these evidentiary tests are prone to giving false negative results due to non-differential measurement error and sample size problems. Also, EMFs may have societally important effects that

are nonetheless truly close to the detection of epidemiology. Finally, an agent may act in an "on/off" fashion and would not produce a steadily increased effect. These patterns of evidence therefore lowered confidence some, but not a lot.

13 CONCLUSIONS

Having examined and discussed each of the health endpoints mentioned above in a separate chapter in the main document, the three DHS reviewers each assigned their best judgment IARC classification and degree of certainty (as a number between 0 and 100). These determinations are summarized in Table V. Column 1 displays the condition considered. Column 2 identifies the reviewer. Column 3 shows the IARC classification in which the number "1" denotes a definite hazard: "2A" a probable hazard, "2B" a possible hazard, and "3" evidence "inadequate" to make a classification. Column 4 displays the pre-agreed-upon phrases for describing zones of certainty. Column 5 shows the ratio of the reviewers imputed posterior odds to the reviewers imputed prior odds (more about this below). In column 6, the reviewers graphed their best-judgment degree of certainty as an "x" and indicated their uncertainty with a shaded bar on either side of that best judgment.

To provide an illustration, this method has been applied to two non-EMF examples in the first two rows. In row 1, Reviewer 2 has indicated that air pollution is a definite causal trigger of asthma attacks and that he is virtually certain of this. In row 2 he shows that he strongly believes that particulate air pollution causes excess deaths. There is relatively little uncertainty around either of these determinations.

Row 3 displays the prior degree of certainty that there would be epidemiologically detectable effects when comparing disease rates among persons exposed to EMFs at or above the 95th percentile of US residential levels to rates at or below the 1st percentile residential exposure. These prior degrees of certainty range from 5 to 12 on a scale from 0 to 100.

Column 5 is labeled "IRL" for "imputed relative likelihood." If the degree of certainty is converted to a probability scale (0–1.0) and, in turn, if one converted the probability to odds (probability/(1–probability)) the imputed prior odds can be compared to analogously calculated imputed posterior odds. One would base these on the "best judgment" posterior degrees of certainty graphed in Table V. The resulting "imputed relative likelihoods" provide some indication of how much the overall pattern of evidence in biophysics, mechanistic, animal pathology, and epidemiological streams of evidence have combined to move the reviewers from their respective starting degrees of certainty. For example, with regard to air pollution triggering asthma attacks, the existing evidence has caused Reviewer 2 to move 900-fold from his prior, while the childhood leukemia evidence has moved him 22-fold*. Royall (Royall, 1997) has suggested anchoring the interpretation of such relative likelihood numbers on the relative likelihoods derived by probability theory from the following hypothetical experiment: Suppose that a reviewer has two urns, one that contains only white balls, the other that contains half white balls and half black balls. He takes one of the two urns at random. To determine which urn he has ended up with, he begins repeatedly withdrawing a ball and then replacing it in the urn (after noting down its color) and mixing up the balls before pulling out yet another ball. If on only one draw he were to find a black ball, he would know that he was dealing with the urn containing 50% black balls. But what is the relative likelihood conveyed by drawing one or more consecutive white balls? Royall demonstrates that drawing 5 white balls in a row conveys a relative likelihood of 32, while drawing 10 consecutive balls conveys a relative likelihood of 1,024. Reviewer 2 views the asthma/air pollution data as being almost as strong as the evidence conveyed by drawing 10 consecutive white balls during the urn experiment, while the childhood leukemia evidence is equivalent to drawing just shy of 5 consecutive white balls.

* Reviewer 2 had a prior of 5 and a posterior for childhood leukemia of 54. The prior odds are 5/95 = 0.0526. The posterior odds are 54/46 = 1.174. The imputed relative likelihood is 1.174/0.0526 = 22.3.

TABLE V. PRIOR AND POSTERIOR DEGREES OF CERTAINTY AND DHS REVIEWERS' APPLICATION OF IARC CLASSIFICATION

CONDITION	REVIE-WER	IARC CLASS	CERTAINTY PHRASE	IRL	DEGREE OF CERTAINTY FOR POLICY ANALYSIS THAT AN AGENT (EMFs) INCREASES DISEASE RISK TO SOME DEGREE
Air Pollution Triggered Asthma Attacks (Example: Not EMF-Related)	2	Human Risk	Virtually Certain	931	0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
Particulate Air Pollution Triggered Deaths (Example: Not EMF-Related)	2	Prob. Risk	Strongly believe	171	0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
Prior Confidence that EMFs Could Cause Epidemiologically Detectable Disease	1	N.A.	Prone not to believe	1	0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
	2		Strongly believe not	1	
	3		Strongly believe not	1	
Childhood Leukemia					0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
	1	1	Strongly believe	140	
	2	2B	Close to dividing line	22	
	3	2A	Prone to believe	17	
Adult Leukemia					0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
	1	1	Prone to believe	29	
	2	2B	Close to dividing line	21	
	3	2B	Close to dividing line	6	
Adult Brain Cancer					0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
	1	2B	Prone to believe	29	
	2	2B	Close to dividing line	20	
	3	2B	Close to dividing line	13	

CONDITION	REVIE-WER	IARC CLASS	CERTAINTY PHRASE	IRL	DEGREE OF CERTAINTY FOR POLICY ANALYSIS THAT AN AGENT (EMFs) INCREASES DISEASE RISK TO SOME DEGREE
Childhood Brain Cancer					0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
	1	3	Close to dividing line	7	
	2	3	Prone not to believe	2	
	3	3	Prone not to believe	3	
Breast Cancer, Female					0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
	1	3	Close to dividing line	7	
	2	3	Prone not to believe	3	
	3	3	Prone not to believe	2	
Breast Cancer, Male					0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
	1	3	Close to dividing line	6	
	2	3	Prone not to believe	12	
	3	3	Prone not to believe	2	
EMF Universal Carcinogen?					0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
	1	3	Strongly believe not	0.4	
	2	3	Strongly believe not	0.5	
	3	3	Strongly believe not	0.2	
Miscarriage					0 5 10 1 5 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
	1	2B	Close to dividing line	9	
	2	2B	Close to dividing line	20	
	3	2B	Close to dividing line	11	
Other Reproductive					0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
	1	3	Strongly believe not	0.4	
	2	3	Strongly believe not	0.8	
	3	3	Strongly believe not	0.2	

CONDITION	REVIE-WER	IARC CLASS	CERTAINTY PHRASE	IRL	DEGREE OF CERTAINTY FOR POLICY ANALYSIS THAT AN AGENT (EMFs) INCREASES DISEASE RISK TO SOME DEGREE
ALS (Lou Gehrig's Disease)					0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
	1	2B	Close to dividing line	9	
	2	2B	Close to dividing line	21	
	3	2B	Close to dividing line	11	
Alzheimer's					0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
	1	3	Close to dividing line	5	
	2	3	Prone not to believe	4	
	3	3	Prone not to believe	2	
Suicide					0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
	1	3	Close to dividing line	6	
	2	3	Close to dividing line	15	
	3	3	Close to dividing line	7	
Heart					0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
	1	3	Close to dividing line	6	
	2	3	Prone not to believe	8	
	3	3	Prone not to believe	3	

14 How different is this evaluation from the NIEHS, NRPB, and IARC findings?

As outlined in Table VI below, there are both common points and significant differences between the EMF Program's evaluation and those carried out at about the same time by the NIEHS (for the Federal EMF-RAPID Program), the NRPB (NRPB, 2001a), (NRPB, 2001b), and the IARC (Note: The NRPB did not use the IARC classification system but expressed their conclusion using common language expressions).

The following table compares these evaluations:

Table VI. A Comparison of DHS Reviewers' Degree of Certainty with That of Other Agencies

HEALTH OUTCOME	NIEHS WORKING GROUP	IARC	NRPB	DHS
Childhood Leukemia	2B*	2B	Possible	2B to 1
Adult Leukemia	2B* (lymphocytic)	Inadequate	Inadequate	2B to 1
Adult Brain Cancer	Inadequate	Inadequate	Inadequate	2B
Miscarriage	Inadequate	Not considered	Not considered	2B
ALS	Inadequate	Not considered	Possible but perhaps due to shocks	2B
Childhood Brain Cancer, Breast Cancers, Other Reproductive, Alzheimer's, Suicide, Sudden Cardiac Death, Sensitivity	Inadequate	Inadequate or not considered	No for Parkinson's Disease, Inadequate for Alzheimer's, Other endpoints not yet considered	Inadequate

It is clear from Table VI that, when applying the IARC guidelines, the DHS reviewers agreed with IARC and NIEHS reviewers that in many cases (e.g., childhood brain cancer and male and female breast cancer) the evidence would be classified by IARC as inadequate to reach a conclusion. One of the DHS reviewers agreed with the IARC and NIEHS on childhood leukemia. Two of the reviewers agree with NIEHS, but not with IARC, on adult leukemia. All three reviewers agreed with NRPB that EMF was a "possible" cause of ALS. Otherwise, the DHS reviewers regard the EMFs association more likely to be causal than NRPB, IARC, or NIEHS did.

It should be noted that all of the review panels thought that the childhood leukemia epidemiology warranted the classification of EMF as a "possible" carcinogen and thus did not agree with the biophysical arguments that EMF physiological effects (and therefore pathological effects) were "impossible."

There is a wide range of opinions in the scientific community as to the probability that EMFs cause health problems. The DHS reviewers provided numerical values for their degrees of confidence that risk of various diseases could be increased to some degree by EMF exposure. Other researchers have rarely packaged their judgments in this way, so it is hard to make comparisons. Judging by one such exercise that the DHS reviewers conducted (Neutra, 2001), reasonable scientists can have different ways of interpreting the data resulting in different degrees of certainty.

* Although the majority of scientists assembled to prepare the NIEHS Working Group Report voted for a "possible 2B" classification for these cancers, the lay person's summary submitted by the Director of NIEHS to Congress stated: "ELF-EMF exposure cannot be recognized as entirely safe because of weak scientific evidence that exposure may pose a leukemia hazard." (Final Report NIH Publication 99-4493, May 1999)

The three DHS reviewers have been active in the EMF field for more than a decade and are familiar with the opinions and arguments used by the scientists in scientific meetings. Since Reviewer 1 was part of the IARC-EMF review panel and all three reviewers had some participation in the earlier parts of the NIEHS process, they also have some understanding of the process by which selected panels of these individuals arrived at a group determination about EMFs. The reviewers think there are at least two relevant differences between their process and the usual procedures followed by the other groups.

First, the DHS Guidelines require that they consider the inherent tendency of the several streams of evidence to either miss a true effect, or falsely "indict" a putative causal agent. The weight given to those streams of evidence was influenced by this consideration. The standard guidelines involve discussions of whether the adjectives "limited" or "sufficient" best fit the pattern observed in a stream of evidence, and depending on the decision one makes, simple guidelines of how combinations of "limited" and "sufficient" streams of evidence influence whether a "possible," "probable," or "definite" causal status is assigned. While the DHS Guidelines allow null results of animal pathology studies using one ingredient of a mixture to get little weight, the IARC rules involve a simple combination of binary judgments about the animal and epidemiological evidence. The way the DHS reviewers used the Guidelines meant that they did not let the primarily null results from the mechanistic and animal pathology streams of evidence decrease their certainty as much as seems to be the case for reviewers in other panels. The reasons for this have been explained above. Having been less deterred by the null mechanistic and animal pathology, they were also less prone to invoke unspecified confounders and bias as an explanation for the persistent, if not homogeneous, epidemiological findings for certain health endpoints.

The other reason for the discrepancies in the DHS reviewers' IARC classification choices can be traced to differences in the procedures for combining the scientists' judgments. They found several striking differences between the IARC and this evaluation processes:

- The Panel's Composition. The EMF Program's review was carried out by the EMF Program's scientific staff and not by a large panel of experts outside the agency. An outside panel, however, evaluated the document. One could criticize the DHS panel as being too small and not diverse enough, but this is standard procedure for California government agencies. The IARC followed its usual practice of convening outside experts to write drafts, discuss the drafts, and turn them over to staff to finalize. Given the spread of the scientific opinions on the EMF issue, it is safe to say that the outcome of any review is a strong function of the working group members' belief before the review takes place. (The DHS reviewers have striven to make this transparent through the elicitation of the prior beliefs and the "pro and con" discussion.) Two unbiased ways to assemble a working group would be by random selection out of a pool of "qualified" individuals or through a conscious effort to include balanced numbers of individuals known to have opposite points of view. In the first case, the definition of "qualified" could influence the verdict of any sample, and sampling variability could yield a mix of opinions that would vary from sample to sample so that different working groups could reach different conclusions. The second procedure could be an excellent solution, if the evaluation were carried out through extensive debates and discussions, with a shared desire to come to a consensus opinion irrespective of its potential social and economic consequences. This was the original approach used by IARC (Tomatis, private communication). However, the pressure to conclude the evaluation within a short period of time led to abandoning the discussion format in favor of the voting system. This leads to the next important difference.
- The Time Element: The meeting to draft the IARC-EMF monograph (June, 2001) lasted five and a half days. The vast majority of the plenary session time was dedicated to reviewing the draft chapters prepared ahead of time by designated committee members with maybe 10% of the time allowed for discussion of the rationale for reaching conclusions. Whenever a paragraph precipitated a controversial discussion, a common way out was to propose the deletion of the offending paragraph, a proposal that the time-pressured working group members were usually glad to adopt. In contrast to this process, the DHS reviewers spent innumerable hours and days, over a period of years and in consultation with independent consultants, to explain their inferences and resolve or clarify their differences.
- The Format of the Conclusion: IARC aims for a consensus conclusion. Members with more extreme views are strongly encouraged to converge on a middle of the road conclusion. In the California evaluation, if consensus could not be reached (as was the case for some endpoints), each member was allowed to express his or her personal belief. Although two of the DHS reviewers were subordinate to the third, substantial differences remained for some endpoints and are openly revealed in this evaluation.
- IARC's Voting System: The members of the working group were asked to vote separately on animal and human evidence. Although a sizable

minority of the working group believed that there was limited animal evidence indicating a possible cancer risk, their opinion was not carried past that point of the process. Since the majority regarded the animal evidence as "inadequate," when the final vote on the overall evaluation was taken, the option posed to the working group's members were the majority positions, that is, that animal evidence was inadequate and epidemiological evidence for childhood leukemia was limited. According to the guidelines, these two majority positions resulted automatically in a Group 2B classification and Class 2A or Class 1 were not even considered as options to vote on, even if individual reviewers, such as Reviewer 1, might have so voted. The published monograph does not document that the minority view had in fact a higher degree of certainty of the EMF risk than the majority view.

Somewhat similar considerations apply to the NIEHS evaluation. Although the whole process lasted eighteen months, the decision was reached over the course of a week-long meeting, followed by a vote. This meeting was preceded by a series of workshops including discussions and presentations, but not all members of the working group participated in the workshops, and most of the workshop participants were not members of the working group. Therefore, the final conclusion was still the result of a few days intensive meeting, during which much of the time was devoted to revising and finalizing the wording of the final report rather than to writing about points of controversy. The working group report did document the vote count.

Apart from procedural differences, there are also philosophical differences between the various review panels. For example, with regard to adult leukemia, the IARC's evaluation differs from the NIEHS and the California evaluation because of the way epidemiological evidence was considered. Almost all the evidence on adult leukemia comes from occupational studies. The Epidemiology subgroup at the IARC meeting regarded most of these studies as being of poor quality, with within- and between-study inconsistencies. Most of the evaluation centered on the most recent large studies (Sahl, Kelsh & Greenland, 1993), (Savitz & Loomis, 1995), and (Theriault et al., 1994), which contradicted each other. The DHS reviewers' evaluation considered the whole body of studies, residential and occupational. While they acknowledge that many of the studies have limitations, neither they, nor the IARC reviewers, have identified fatal flaws. For example, there is no evidence to suggest that the use of crude exposure assessment surrogates, while virtually certain to influence the quantitative estimate of risk and to frustrate any attempt to explore the dose-response relationship, introduced an upward bias in the reported association. On the contrary, the limitations of the studies may well be responsible for the inconsistencies between them. And while these inconsistencies do exist, they are not as common as the IARC evaluation may suggest. The Kheifets (1997) meta-analysis concludes that the body of epidemiological evidence shows a slight but statistically significant increase in risk. From a binary outcome standpoint, the studies with a relative risk estimate >1 are more than twice as numerous as those with a $RR \leq 1$.

Nonetheless, where the DHS and other reviewer panels agreed to assign a "possible" carcinogen label to an EMF/disease association, it is not easy to infer if there would be agreement on a degree of certainty. According to Dr. Rice, Chief of IARC's Carcinogen Identification and Evaluation Unit (personal communication to Vincent DelPizzo), "If IARC were to say that an exposure is in Group 2A, probably carcinogenic to humans, that would mean that the evidence is just a little short of certainty that the exposure in question has actually caused human cancer. . . Group 2B is the lowest level of identifiable carcinogenic hazard in the IARC system."

Finally, it must be remembered that in DHS's EMF Program, policy recommendations were addressed separately from the risk evaluation. In some other cases, evaluations are part and parcel of a policy recommendation (they may include regulatory recommendations in the conclusion). This may make them more conservative, as it seems to be the case with IARC: "….the IARC Monographs system of carcinogenic hazard evaluations is deliberately a very conservative one. There are many carcinogenic hazards in the human environment that are very real indeed, and control of exposures to those hazards is extremely important for public health. To accomplish this, it is necessary that carcinogenic hazards be correctly identified. We must avoid misdirecting public attention to any exposure of any kind that may be perceived as a hazard, but in fact is a misplaced concern." (Dr. Jerry Rice in a letter to Vincent DelPizzo, Aug. 10, 2001). The cover letter to the NIEHS report to congress concluded with a recommendation for only "passive regulatory action" (NIEHS, 1999). The DHS's three reviewers have packaged their differing degrees of confidence about causality in a way that can be used in the decision analytic models prepared for the program. DHS has pointed out that the policy implications of this range of confidences depends on the policy framework of the decision maker: non-interventionist, utilitarian, virtual-certainty-required, or social justice. The public regulatory process will determine which one or which mixture of these frameworks will apply to govern policy. Thus the DHS risk evaluation is packaged to facilitate decision making but separates risk assessment from risk management. The fact that a reviewer may feel very certain that EMF is a risk factor for a particular disease does not imply that he or she advocates exposure mitigation.

In summary, the differences between the DHS reviewers' judgments and those of other reviewers are partly due to differences in procedure and terminology and partly due to the way those three reviewers weighed the several streams of evidence.

15 Differences Between DHS Reviewers

As noted above, the three DHS reviewers were not able to reach a consensus on all health endpoints. In this section, they explain the reasons behind their respective judgments.

15.1 Reviewer 1 (DelPizzo)

In almost all cases, Reviewer 1's posterior degree of confidence is higher than that of the other two reviewers. There are several reasons for this difference.

a) Different priors—the reviewer is generally more suspicious of man-made environmental pollutants, which have no place in the evolution process.

b) Reliance on the sign test—this reviewer has put much weight in the sign test, a simple, dichotomous test, which measures the probability of several studies erroneously reporting the existence of a risk while no risk truly exists. In many cases the test finds that this probability is extremely small, that is, the results are unlikely to be erroneous. In the reviewer's opinion, this test is particularly suitable to answer the simple question, is there a risk or not? rather than asking what the relative risk is. The results of this test are not changed if the outcome of one or more studies are partly due to bias. Some worst-case scenarios, assuming extraordinary coincidences of chance and bias acting simultaneously in the same direction, do weaken the evidence, but when a condition has been studied by many different investigators, these scenarios do not reduce Reviewer 1's belief by much.

c) Weight given to empirical results—Reviewer 1's prior was limited by the intuitive belief that the energy associated with environmental EMFs is so small that, even if these fields are potentially disruptive, the amount of disruption is insufficient to cause a biological effect. Once Reviewer 1 examined the results of in vivo and in vitro research on EMF exposure, however, he became convinced that biological EFFECTS (as distinct from PATHOLOGY) can result from exposure to levels below those which conventional knowledge considers necessary. That is, if one equates "energy" to "dose," exposure to environmental fields may be regarded as a non-negligible dose. Thus, the argument that kept Reviewer 1's prior low disappears and the possibility of a hazard, when repeatedly reported by independent epidemiological studies, becomes more credible.

15.2 Reviewer 2 (Neutra)

The fact that EMFs are the only agent that this reviewer has encountered for which there are theoretical arguments that no physiological, much less pathological, effect could be possible, did decrease Reviewer 2's prior somewhat. But physics applied to simplified models of biology were not convincing enough to make this prior credibility vanishingly small. This reviewer noted biological effects in mechanistic experiments in the thousands of mG but accepted the arguments that these were probably not relevant to effects below 100 mG. The few experiments that claimed to show an effect below 100 mG (the chick embryo studies and the confirmatory studies of Liburdy's melatonin studies) were considered highly worthy of further study, but not robust enough or free enough of alternative explanations at this point to cancel out the modest initial doubts about the energetic feasibility of residential EMFs to produce biological effects. The animal pathology studies have convinced Reviewer 2 that very-high-intensity pure 60 Hz or 50 Hz sinusoidal magnetic fields do not have a strong enough effect to produce consistent pathological effects in small numbers of the species and strains of animals selected for study. If these species of animals were to respond as humans are described to have done in the epidemiology, this was a predictable result even if pure sinusoidal 60 Hz fields were the active ingredient of the EMF mixture. Humans exposed to hundreds of mG, like electric train engineers, when compared to persons with 24-hour average exposures around 1 mG do not show relative risks consistently above 1.00 much less very high relative risks. Why would animals be expected to do so? Moreover, pure sinusoidal fields may not be a bioactive ingredient of the mixture, and the animal species chosen may not be appropriate models for humans. Reviewer 2 believes that the animal bioassay stream of evidence in this case is thus triply vulnerable to missing a true effect, and the null results do not reduce his confidence in an EMF effect much. The fact that there are epidemiological associations with several different cancer types and with other diseases that have different known risk factors does increase confidence somewhat but, without mechanistic reasons, not a great deal. Any changes from the prior were due to epidemiological evidence. Large studies likely to be free of selection bias carried a lot of weight. Many studies of different design and in different locations showing similar results also carried substantial weight, although Reviewer 2 only interpreted the sign test to indicate whether a meta-analytic or pooled association came from just a few large studies, or from a rather consistent pattern of result from many studies. Reviewer 2 did not think that any of the specific

candidate confounders or biases that had been proposed to date for explaining away the epidemiology had convincing evidence to support it. The fact that most of the associations are not much above the resolving power of epidemiological studies left open the possibility of unspecified combinations of bias, confounding, and chance having produced these associations. This kept Reviewer 2 from having an updated degree of confidence above the certainty zone of "close to the dividing line between believing and not believing" that EMFs increase the risk to some degree.

15.3 Reviewer 3 (Lee)

Reviewer 3 mainly used the human epidemiological evidence to form a posterior degree of confidence. The large number of studies showing consistent results across different study designs, study populations, and exposure assessments, as well as large, well-conducted studies with adequate power to address confounding, bias, dose response, and effects among subgroups contributed strongly in updating the prior degree of confidence. The association of EMF with several types of disease and experimental and animal evidence were minor contributions to the updating process. Specificity, visibility, analogy, and, in general, temporality did not contribute much to the posterior degree of confidence.

16 How The Degrees of Confidence and Range of Uncertainty Could be Used in Policy Analyses

Community and stakeholder policy decisions usually are made from one or more of the following ethical perspectives: "non-interference," which emphasizes individual choice and rights free from the infringement of others and of government; "social justice," which emphasizes the protection of the weak, and rights and duties; "virtual-certainty-required," where protective action is only taken when the vast majority of scientists are virtually certain that there is a problem; and the "utilitarian perspective," which emphasizes results and the most good for the most people at the least cost. Each perspective would have somewhat different requirements for the degree of confidence of causality before initiating action.

The "non-interference" perspective seeks to avoid regulatory impingement and taxes and tends to favor "right to know" warnings and voluntary solutions to problems, regardless of the degree of confidence. The "virtual-certainty-required" framework would tend to require a high degree of confidence with narrow uncertainty bounds on the part of most scientists and a high probability of harm from exposure before acting on an environmental hazard. Indeed, this perspective would favor risk-assessment methods having few false positives, even at the cost of false negatives.

The "social justice" perspective seeks to avoid even the possibility of risk, particularly if the risk and the benefit are imposed on different parties. This perspective would tend to advocate protective action at lower degrees of confidence, wider uncertainties, and lower absolute probabilities of harm given exposure. It would favor risk-assessment approaches with few false negatives, even in the face of false positives. It would focus on the added lifetime risk to the most highly exposed.

The "utilitarian cost/benefit" perspective would evaluate the policy implications of the best estimate of the degree of confidence but would explore the consequences of the lower and upper bounds of the confidence that a hazard exists. It would focus on the burden of societal disease that could be avoided by EMF mitigation. Depending on the relative prevalence of stakeholders who suffer, respectively, from false positives and false negatives, the utilitarian perspective would develop a preference for risk-assessment methodologies. The reviewers would propose that the policy integration document discuss the implications for policy arising from the range of best estimates among the three reviewers and the range of uncertainties expressed. It should also discuss where the three DHS reviewers' degrees of confidence lie in the spectrum of scientific opinion.

17 Evidence of Risk Relevant for Policymakers Mindful of Environmental Justice Issues

It is sometimes alleged that lower SES subjects are more likely to live in areas with stronger environmental EMFs. Salzberg et al. (Salzberg, Farish & DelPizzo, 1992) first explored this hypothesis and found only weak support for it. Bracken et al. (Bracken et al., 1998) reported a strong correlation between some SES indicators (women's occupations, house values) and the very high-current configuration (VHCC) wire code configuration. Hatch (Hatch et al., 2000) found no such association. Two very large data sets collected in the San Francisco Bay Area as part of the study by Lee et al. (Lee et al., 2002) found no evidence of an association between family income and measured EMF exposure. However, there was a weak association between low SES and wire code (Hristova et al., 1997). In a geographic information system (GIS) study as part of the power grid policy project, English et al. (http://www.dhs.ca.gov/ehib/ emf/ pdf/ AppendixG-GIS.PDF) examined the ethnic and income characteristics of census blocks within 500 feet of transmission lines. The proportion of black and Hispanic residents in these corridors was lower than the state average proportion. Zaffanella and Hooper (Zaffanella & Hooper, 2000) found somewhat higher magnetic fields in schools with students of lower socioeconomic status. In summary, the evidence to support the contention that the EMF exposure,

if real, disproportionately affects low SES subjects is not very strong, but there is some suggestive data that decision makers may consider when evaluating policy options.

18 THE EMF MIXTURE

A careful assessment of the electricity-related exposures from power lines, appliances, and occupations would reveal what amounts to a complex mixture including electrical and magnetic fields with their respective frequency, polarization, etc. The reviewers will call these the "aspects" of the mixture.

Each aspect varies from instant to instant to form a time-series of intensities, which can be summarized as a single number by various summary "exposure metrics," which may be more or less biologically active. For example, the exposure metric of ionizing radiation that best predicts biological effects is the simple integral of the exposure-time series. The exposure metric that best predicts the effect of an antibiotic might be the integral of blood levels above some threshold. Other electricity-related correlates of proximity to power lines, internal wiring, and appliances are not part of the fields at all, but might be correlated with them. These include electrically charged and "sticky" air pollution particles; contact currents from stray currents, from plumbing and in the earth, and intermittent shocks. The reviewers will call these the "ingredients" of the mixture.

What aspects, ingredients, or exposure metrics, if any, should we be considering in this risk evaluation?

For a number of years, some researchers believed that if the risk increase were truly due to some component of the EMF mixture then this component must be something captured by the exposure-assessment surrogate known as "wire coding," consisting of classifying residences based on their proximity to visible power lines and on the type of these power lines. Recent new data and reanalysis of old data (Linet et al., 1997), (Greenland et al., 2000) appear to have disposed of this hypothesis convincingly. They have shown that risk is more consistently correlated to measured or calculated TWA magnetic field than to wire coding classification.

This does not mean that the TWA—measured by surrogates such as point-in-time or "spot" measurements, calculations using engineering models and historical line current loads and job exposure matrices—is necessarily the true causal agent. The units, mG or μT, that measure the magnetic field's TWA do not describe the magnetic field (and much less the electric field associated with it) any more than the units marked on the volume dial on a stereo system fully describe the sound coming out of the speakers.

Nevertheless, although the reviewers cannot definitely "rule in" the component(s) of interest, they can rule out some aspects of the fields that are not correlated with TWA field strength. A detailed discussion of this issue can be found in Neutra and DelPizzo (2001). Here, the reviewers include Table VII adapted from that paper, pointing out which of the more commonly proposed metrics are indeed correlated with TWA (indicated by a "**U**") and those which are not (indicated by "No"):

TABLE VII. CORRELATION OR ABSENCE OF CORRELATION BETWEEN EXPOSURE METRICS AND EXPOSURE-ASSESSMENT SURROGATES

EXPOSURE METRIC TO 30-300 HZ MAGNETIC FIELDS	HIGH WIRE CODE	HIGH MEASURED FIELD	HEALTH ENDPOINT	REFERENCE
(1) TWA	**U**	**U**	**U**	many
(2) Length of time with constant field above a threshold	**U**	**U**		
(3) Repeated periods of elevated exposure	**U**	**U**	**U**	(Feychting, Forssen & Floderus, 1997), (Feychting, Pedersen & Svedberg, 1998b). (Lee & McLoed, 1998)
(4) Third harmonic	**U**	?	?	(Kaune, 1994b)
(5) Resonance with static field	No	No	?	(Kaune, 1994b), (Bowman, 1995)
(6) Time above a threshold	**U**	**U**	?	(von Winterfeldt & et. al., 2001)
(7) Polarization	?	?	?	(Burch et al., 2000)
(8) Transients	No	No		(Preece et al., 1999)
(9) Maximum daily exposure	**U**	**U**	**U**	(Li et al., 2002), (Lee et al., 2002)
(10) Average change between measurements	**U**	**U**	**U**	(Lee et al., 2002)
(11) Electric field	Not inside home	Not inside home	?	(Miller et al., 1996), (Coghill et al., 1996)

This table allows the reviewers, at least, to cast doubt on two metrics that are supported by mechanistic arguments, but not (or at least not consistently) by empirical data. These are 1) magnetic field transient, which can induce strong, if brief, electrical currents in the body, and 2) resonance conditions, which may facilitate energy transfer from the field to the living organism.

The table also emphasizes the difficulty of testing the hypothesis of an EMF risk by conducting experimental studies. Studies using an exposure apparatus that delivers an appropriate TWA (but not an appropriate exposure to a hypothetical aspect, ingredient, or exposure metric found in residential or occupational environments) are liable to produce false-negative results. Or they may produce positive results suggesting dose-response relationships different from those that may result from environmental fields.

Reducing TWA exposure will reduce exposure to several other metrics and reduce any risk from TWA or the exposure metrics that are changed with it. However, this is a sufficient but not necessary condition: if TWA is not by itself the causal factor and if we could identify and remove from the EMF mixture the component directly causally associated with the health endpoint, a subject could still be exposed to high TWA and not be at risk. Also, because the correlation coefficient between TWA and these other components of the field are modest to moderate, reducing TWA exposure would not reduce the risk proportionally to the decrease in the average field strength.

The following table compares the values of the magnetic field strength, measured by direct personal measurement or by environmental monitoring (spot or 24-hour measurements). Note that these are not data collected on the same sample, but general information gleaned from the literature (Zaffanella & Kalton, 1998), (Lee et al., 2002) and mathematical modeling.

TABLE VIII COMPARISON OF THE VALUES OF THE MAGNETIC FIELD (MG) STRENGTH MEASURED BY DIRECT PERSONAL MEASUREMENT WITH ENVIRONMENTAL MEASUREMENTS

PERCENTILE POINT OF EACH TYPE OF MEASUREMENT	TWA PERSONAL FIELD	AVERAGE SPOT HOME MEASUREMENT	MEDIAN SPOT HOME MEASUREMENT	MEDIAN 24-HOUR HOME FIELD
99	5.5	6.6	5.8	5.5
95	3.2	3	2.6	2.6
90	2.4	2.1	1.7	1.8
75	1.5	1.1	1	1
50	0.9	0.6	0.5	0.5

The personal TWA is generally higher than the environmental levels, reflecting the contribution that occasional close proximity to localized sources (appliances, wall wires, buried cables) makes to the average personal exposure. However, at the upper end of the distribution, this difference is minimal or non-existent, reflecting the fact that exposure to localized sources is common to all subjects. These localized sources contribute a few tenths of a mG to the personal 24-hour average (TWA).

What determines the "exposed" status of a subject in epidemiological studies (generally defined as a TWA above 2–4 mG) is usually the background environmental exposure, and that is contributed largely by home exposure (where people spend the most time). Certain occupations are an exception to this generalization because work-time exposure is so much higher than home exposure. According to Zaffanella's "1000 homes study" (Zaffanella, 1998), these background fields are due, with almost equal frequency, to proximate power lines and to grounding system fields.

Of course, this conclusion about background fields will change drastically if future research confirms the hypothesis-generating data by Lee (Lee et al., 2002) and Li (Li et al., 2002), indicating that, at least for spontaneous abortion (SAB), the true risk factor is the maximum daily exposure above 14 mG or the average field change between measurements. If maximum exposure, or one very strongly correlated to it, is the appropriate metric, then sources of localized fields (appliances, home wiring) become more important than power lines and ground currents because the latter seldom produce fields of the intensity implicated by the Lee and Li studies.

An additional difficulty that arises in this case is that personal measurements taken at the hip, as is common practice, may introduce errors that are large compared to the instrument error. This is because the field produced by a localized source shows significant variation based on which anatomical site is measured (DelPizzo, 1993),

even though some sources like power lines outside the house may produce a field at locations like the eye and the hip that are virtually identical. We also have no clear evidence by which to determine if the EMFs interact with biological systems at specific target organs. For example, there is some evidence that birds perceive geographic variations of the earth's magnetic field by means of their eyes (Graves, 1981). On the other hand, EMFs might act directly on cells in the marrow or in the uterus. Personal measurements taken at the hip might miss some exposures to the eye, but not exposures to the uterus.

It must be stressed that, although the Li (2002) and Lee (2002) studies are recent, good-quality studies with similar results, they have not yet been replicated. While meriting attention, they do not negate the wealth of data associating 24-hour average field to risk of other diseases.

19 Potential Annual numbers of Deaths Attributable to EMFs

Two recent review articles calculated the proportion of all childhood leukemia cases that might be attributed to the rare highest residential EMF exposures. This was estimated to be around 3%. With about 100 childhood leukemia deaths per year, this would translate to about 3 deaths in California per year attributable to EMFs. The evidence does not permit similar direct calculations for the other reviewed conditions. However, suppose that only 1% of the conditions that were considered in this evaluation (minus those that the three reviewers "strongly believed" were not caused by EMFs) could be attributed to EMF exposure. The numbers of attributable cases could still be in the hundreds per year and comparable to the theoretical burden of ill health that has motivated other environmental regulation (di Bartolomeis, 1994). The annual California deaths from each of these conditions are shown in Table IX. The reader can apply 1% to these numbers to verify the assertion in the previous sentence.

Table IX. 1998 Yearly California Deaths (some fraction of which might be affected by EMFs) *

AGE GROUP	CHILD LEUK.	ADULT LEUK.	CHILD BRAIN	ADULT BRAIN	MALE BREAST	FEMALE BREAST	SPONT. ABORT.+	ALS	ALZ-HEIMER	SUICIDE	ACUTE M.I.
0-19	99	0	79	0	0	0	11,000	0	0	171	2
29 Plus	0	1888	0	1294	30	4095	49,000	434	320	3044	17,236

* From http://www.ehdp.com/vn/ro/av/cau1/eg1/index.htm

+ Note: many would not consider spontaneous abortion as serious as the death of a child or adult.

20 Potential Added Lifetime Risk from High Exposure

Since epidemiology is a blunt research instrument, the theoretical lifetime individual risk that derives from any agent that has an epidemiologically detectable effect will be automatically greater than the lifetime risk of 1/100,000 that triggers many regulatory processes. This means most of the epidemiological associations examined in this document could clearly be of regulatory concern if real.

That being said, with the exception of miscarriage, the theoretical lifetime risks from the highest EMF exposures are such that, depending on the disease and assuming relative risks ranging from 1.2 to 2.0, 93% to 99.9% of even highly exposed individuals would escape contracting the non-miscarriage health conditions studied.

These insights are illustrated in Table X below.

TABLE X. ADDED LIFETIME RISK IMPLIED BY RELATIVE RISKS OF 1.2 OR 2.0 FOR RARE AND COMMON DISEASES

ANNUAL INCIDENCE	DISEASES IN CATEGORY	ADDED ANNUAL RISK FROM: RR =1.2; RR= 2.0	ADDED LIFETIME RISK FROM: RR = 1.2, RR = 2.0	LIFETIME CHANCE OF ESCAPING DISEASE AFTER EXPOSURE
1/100,000	ALS, Male Breast Cancer	0.2/100,000 ; 1/100,000	1.4/10,000; 7/10,000	99.99%; 99.93%
5/100,000	Child Leukemia	1/100,000; 5/100,000	2/10,000; 10/10,000	99.98%; 99.9%
10/100,000	Suicide, Adult Brain, & Leuk.	2/100,000; 10/100,000	14/10,000; 70/10,000	99.9%; 98.3%
100/100,000	Acute Myocardial Infarction	20/100,000; 100/100,000	1.4%; 6.8%	98.6%; 93.2%
1%	Alzheimer's	0.2%; 1%	NA (late onset)	NA
10%	Miscarriage	2%; 10%	NA (occurs during pregnancy)	NA

Note: RR = risk ratio; NA = not applicable

Two new epidemiology studies (Li et al., 2002), (Lee et al., 2002) suggest that a substantial proportion of miscarriages might be caused by EMFs. Miscarriages are common in any case (about 10 out of 100 pregnancies) and the theoretical added risk for an EMF-exposed pregnant woman may be an additional 10 out of 100 pregnancies according to these two studies. If true, this could clearly be of personal and regulatory concern. However, the type of EMF exposure implicated by the new epidemiological studies (short, very high exposures) probably come primarily from being very close to appliances and indoor wiring, and only rarely from power lines. Seventy-five percent of the women in the studies had at least one of these exposures during a day, and even one exposure a day, if typically experienced during pregnancy, seemed to increase the risk of miscarriage. Nonetheless, the vast majority of pregnant women with such exposures did NOT miscarry.

21 POLICY-RELEVANT AREAS FOR FURTHER RESEARCH

One of the major impediments to evaluating the potential bioactivity of a complex mixture is identifying the bioactive components of that mixture. This usually requires finding some kind of bioassay with which to assess the mixture and then successive fractions of it. While some epidemiologists have attempted to evaluate the effects of different aspects of the EMF mixture and some exposure analysts have attempted to characterize the occurrence and intercorrelation of its aspects, important policy-relevant questions still remain.

Experimentalists have rarely used the mixture as it occurs in real life and have focused instead on one or the other aspect of the mixture, usually pure sinusoidal 60 Hz fields at intensities far above those found in residential or blue collar occupational environments. Deeply ingrained experimental research styles and an orientation to explaining mechanisms rather than describing phenomena has meant that investigator-initiated research and even programs that attempted to guide research have rarely been characterized by progressively refined descriptions of dose-response relationships to produce stronger bioeffects.

This has been compounded by the expectation of a quick resolution of the question by those who fund research, as was the case with the New York State program of the mid-1980s, the current California Program, and the recent five year federal EMF–RAPID program. As was discovered after President Nixon's "War on Cancer" in the early 1970s, research progresses slowly and in successive multi-year research cycles, with the results of each cycle governing the direction of the next. It would not be surprising if it took four more five-year research cycles to clarify the EMF issue.

This means that if one were serious about clarifying this issue there would need to be a long-term commitment to steady research funding and funding for intermittent assessments of the state of the science and research directions. Most research peer review groups would favor research where a clear bioeffect was present and credible alternative mechanisms were being explored. Those situations tend to have

a high yield of early definitive results, and such results lead to continued research funding, publications, and research career advancement. The EMF area does not fit this description and from this perspective would receive a low priority for funding from the usual peer review study sections. Indeed, prominent researchers who doubt that there are any bioeffects, much less epidemiological effects, from the residential and occupational EMF mixture, feel there is nothing to find and have recommended that no more funding for this area be provided (Park, 1992).

Clearly the three DHS reviewers disagree with the assessment of the evidence to date and see a number of research areas which are worth pursuing that could influence and focus exposure avoidance strategies, if any. The cost effectiveness of further research has been a topic of the program's policy analysis and will be discussed at greater length in our policy integration document. The cost/benefit analysis of EMF research suggests that there is so much at stake in choosing between "expensive," "inexpensive," and "no mitigation" that more research funding can be easily justified. (http://www.dhs.ca.gov/ehib/emf/pdf/Chapter09-ValueofInformation.pdf)

The highest initial priorities for the reviewers would be to carry out exposure studies in residential settings and the workplace to see if purported aspects of the EMF mixture that would require different mitigation strategies are correlated with magnetic field exposure and could therefore explain their apparent effect. Such aspects include sudden exposures to the 60 Hz fields, such as micro-shocks, stray ground currents, and charged air pollutants. Such exposure studies would make it possible to reanalyze some of the existing worker cohorts to determine if these aspects are associated with diseases.

Rather than further pursuing new studies of rare diseases with long incubation periods, further studies of the more common conditions in which EMFs might have shorter induction periods, such as spontaneous abortion, acute myocardial infarction, and suicide should be given priority. These would be more relevant to a utilitarian policymaker.

On the experimental front, the reviewers suggest giving priority to finding reliable bioeffects below 100 mG and to carefully exploring dose-response relationships and then mechanisms. The balance between investigator-initiated and programmed research, as well as the guidelines that will be used for interpreting results, need to be carefully considered.