EPIDEMIOLOGY IN HAZARD AND RISK ASSESSMENT

A Partial Review of the "London Principles",
With Recommendations for Revisions and
Additions, by an Expert Panel

Based on a Review of Studies
of Possible Relationship Between
Induced Abortion and Breast Cancer

November 1999

Federal Focus, Inc.®
Washington, DC, USA

TABLE OF CONTENTS

INTRODUCTION

This report is the latest step in a multi-phase project to develop generally agreed upon principles or other methods for evaluating epidemiologic data for use in human health hazard identification and risk assessment, particularly when such hazard identification or risk assessment is intended to form a basis for regulatory action or government dissemination of health information.

The first two phases resulted in the August 1996 publication by Federal Focus of PRINCIPLES FOR EVALUATING EPIDEMIOLOGIC DATA IN REGULATORY RISK ASSESSMENT.[1] Those principles, which we refer to as the "London Principles", were drafted by a panel of eighteen experts in public health and medicine in London, England in October 1995. The London effort followed a preliminary conference by a similar panel in 1994.[2] The London Principles were divided into basically two parts: A set of principles for evaluating an epidemiologic report for cause-effect relationship, and a set of

---

[1] ISBN 0-9654148-0-9. A limited number of copies are available from Federal Focus. The Principles and Recommendations portion of the London report is also posted on the Federal Focus website, www.fedfocus.org, and Federal Focus has waived copyright to that portion of the report, as well as those portions of the current report containing the specific recommended revisions and additional principles.

[2] THE ROLE OF EPIDEMIOLOGY IN REGULATORY RISK ASSESSMENT (J. Graham ed., Elsevier Science, B.V 1995, ISBN 0-444-82201-1).

principles for using human and animal data in dose-response evaluation. Additionally, at the request of the London panel, Federal Focus developed and published separately at the same time a companion set of recommendations for implementing the London Principles in governmental hazard and risk assessment guidance.[3]

Subsequently, Federal Focus convened a number of workshops with Federal agency and private sector scientists and risk assessors to obtain their views on the London Principles. While the principles were generally regarded as sound, many participants in those workshops, as well as members of the original London panel, urged Federal Focus to organize one or more "test runs" of the principles, using them to evaluate specific databases and thereby acquire more insight into their practical utility and any need for modifications or enhancements.

---

[3] "Epidemiologic Data in Regulatory Risk Assessments: Recommendations for Implementing the London Principles and for Risk Assessment Guidance" (Federal Focus, Inc. 1996). Copies are available from Federal Focus, Inc.

Federal Focus followed these recommendations, and began reviewing the general composition of databases for a number of potential hazards likely to be of general interest. After spending some time in this review process, and consulting with a number of interested parties, we came to two basic conclusions: First, the "test run" would be more manageable if we concentrated on only one of the two sets of principles -- the cause-effect relationship set.[4] Second, the exercise would be more likely to be perceived as completely objective if we could avoid review of an exposure which was of current regulatory interest, or might be in the near future. We also had to take into consideration the need for a database including a significant number and variety of epidemiologic studies, the desirability of including in the test run one or more recent meta-analyses, and the desirability of having issues of biological plausibility and dose-response related to interpreting the epidemiologic findings. Such considerations led us to select for the test run the body of studies on the possible relationship between induced abortion and breast cancer. That database also had the following characteristics that seem to appear often in the databases for many currently controversial risk issues:

---

[4] This report is titled a "partial review" of the London Principles because it addresses mainly the cause-effect principles and questions (including Principle B-6 on meta-analysis and pooling, which was originally placed in the dose-response set), and does not address the dose-response evaluation principles. Although the title of this report refers to hazard and risk "assessment" for simplicity, there is a more technically accurate distinction between hazard "identification" and risk "assessment", since "assessment" indicates an attempt to quantify risk taking the level of exposure or dose into account, whereas hazard "identification" usually indicates a simple Yes or No determination of causal relationship without regard to level of exposure or dose.

- · apparently inconsistent study results
- · relatively low reported "positive" risk ratios
- · potential for recall/response bias
- · possible confounding by other risk factors
- · a recent and reasonably comprehensive meta-analysis combining statistical analysis and qualitative discussion
- · possible issues regarding coherence between epidemiologic study results and biological findings

An effort was made to assemble a body of epidemiologic studies on the subject that was as comprehensive as feasible. A selection of pertinent animal and human biological studies was added for consideration of factors such as biological plausibility and dose-response. Those materials, as distributed to the panel members in full text (with the exception of translation of only pertinent portions of some foreign-language articles), are listed at the end of this publication. There are 31 epidemiologic study and review articles, and 7 articles or book chapters on animal experiments and human biology. As explained below in the "Description of the Review Process", only a subset of the epidemiologic studies were assigned for detailed review and discussion, while the others were optional.

Consistent with the recommendations of the London panel, Federal Focus recruited a review panel that was multidisciplinary. Four panel members were expert cancer epidemiologists; one was an expert in pharmacology and toxicology[5]; and one was a clinical practitioner in reproductive endocrinology, ob/gyn, and psychiatry. Only one of the epidemiologists had previously worked on the specific association being

---

[5] Dr. Waddell was unable to attend the meeting in Denver, and participated via teleconference in the initial portion of that meeting concerning biological issues. He subsequently reviewed the full transcript of the meeting and submitted additional comments to the rest of the panel. His post-meeting comments are incorporated into the discussion summary as if he had been present. He also contributed to drafting of the "Additional Principles for Hazard and Risk Assessment".

reviewed. The panel members and their affiliations are listed following this Introduction.[6]

As described in the following section on the review process, this review panel met for a full day in Denver, Colorado in July 1998, and engaged in extensive pre-meeting and post-meeting work.

The fundamental purpose of this review exercise was to critique the London Principles. The specific review materials were considered a device to test practical application of the Principles and coax focused thinking on their accuracy, completeness, and practicality. Thus, the panel did not arrive at any overall conclusions on causal relationship for the particular database. Moreover, since only a subset of the literature was reviewed in any detail, this would not have been appropriate. Nevertheless, the discussion summary does reflect opinions by individual panel members on some of the more notable studies which individual readers might wish to consider in conducting their own review and drawing conclusions.[7]

---

[6] One panel member, Dr. Friedenreich, had participated in the Federal Focus 1994 Lansdowne conference, which laid the groundwork for the London conference. Dr. Friedenreich also contributed a paper to the publication resulting from that conference, THE ROLE OF EPIDEMIOLOGY IN REGULATORY RISK ASSESSMENT, *supra:* Friedenreich C, "An evaluation of biases introduced by confounding and imperfect retrospective and prospective exposure assessments", at pp. 43-51. The other three epidemiologists on the Denver panel had been quoted or referenced in the *Science* article by G. Taubes which was reprinted as Appendix G of the London report.

[7] To facilitate candor, it was agreed beforehand that there would be no attribution of views.

This report contains numerous recommendations for revisions in the original hazard identification principles. It also recommends an almost complete revision of the principle covering meta-analysis and pooling, which was initially placed in the dose-response set of the London Principles (B-6). Finally, the principles and questions for hazard and risk assessment are completely new.

It may be necessary to read the Discussion Summary section to discern the rationales for the specific recommended revisions and additions.[8]

All involved in this phase of the project had to work within time and budget constraints, and the work presented in this report, while a best effort within those constraints, surely cannot purport to be definitive. Rather, this report, together with the previous reports, is, in our opinion, best regarded as a work in progress, on which others will doubtless wish to comment. We look forward to this. Federal Focus has previously posted the London Principles on its website (www.fedfocus.org) and will also post the recommended revisions and additional principles from this report. We will also provide a means for interested parties to comment interactively on that site.[9]

---

[8]   The Discussion Summary was prepared by the moderator from a full transcript of the meeting, and underwent limited review and editing by the panel members.

[9]   This project was financed solely by Federal Focus, Inc., and was not supported by any grants or specific contributions.

6

PANEL MEMBERS

Sarah L. Berga, M.D.
Dept. of Obst., Gyn. and Repro. Sciences
Magee-Women's Hospital
The University of Pittsburgh
   School of Medicine
Pittsburgh, Pennsylvania
USA

Janet R. Daling, Ph.D.
Professor of Epidemiology
University of Washington, and
Program in Epidemiology
Fred Hutchinson Cancer Research Center
Seattle, Washington
USA

Christine Friedenreich, Ph.D.
National Health Research Scholar
Division of Epidemiology, Prevention
   and Screening
Alberta Cancer Board, and
Adjunct Assistant Professor
Dept. of Community Health Sciences
University of Calgary
Calgary, Alberta
Canada

Sander Greenland, Dr.P.H.
Professor
Dept. of Epidemiology
UCLA School of Public Health
Topanga, California
USA

David A. Savitz, Ph.D.
Professor and Chair
Department of Epidemiology
School of Public Health
The University of North Carolina
Chapel Hill, North Carolina
USA

William J. Waddell, M.D.
Professor and Chair Emeritus
Department of Pharmacology & Toxicology
School of Medicine
University of Louisville
Louisville, Kentucky
USA

DESCRIPTION OF THE REVIEW PROCESS

Twelve published epidemiologic study reports were selected for more detailed review and discussion for the purpose of considering the utility of the London Principles and any recommendations for revisions. Those articles are referenced in the Publications section of this report. Since the goal of the exercise was not to arrive at conclusions regarding the overall database, there were not rigorous criteria for selection of these twelve studies. The studies were selected largely to provide a range of both positive and negative (or null) findings, diversity in study population origins and study design, and apparent influence on the statistical portion of the meta-analysis presented in the Brind et al. meta-analysis article.

Each of the four epidemiologists on the panel was assigned six of the twelve studies to review and comment on; and each study had two assigned reviewers/lead discussants. Assignments were made by random drawing (with the exception that no one was assigned to review a study to which they had contributed). The epidemiologists reviewed their assigned studies prior to the meeting using a copy of the London Principles and actually checking boxes where called for and noting comments and questions on the forms.

The panel met for a full day in Denver, Colorado during July 1998. The meeting was moderated by William G. Kelly, Jr. of Federal Focus.

The two non-epidemiologists began the meeting with a discussion of the relevance of the experimental animal and human biological data and the biological plausibility of the hypothesis. The four epidemiologists then took the lead in the discussions. The two assigned reviewers for each study presented their comments and their experience in applying the London Principles to the study, and then there was plenary discussion, with particular emphasis on what was learned about the accuracy and utility of the London Principles and what, if anything, should be done to revise or augment the Principles.

After all of the review presentations and related discussion, there was further discussion and agreement on the intended work product and post-meeting work assignments. The four epidemiologists divided the London Principles into three sets (A-1 to A-3, A-4 to A-6, and B-6), and assigned lead

responsibility for developing specific recommendations for revisions or additions. The two non-epidemiologists, along with the moderator, took lead responsibility for developing the recommended additional principles for hazard and risk assessment. The moderator also prepared the discussion summary from the transcript which appears in this report and circulated it to the panel for use in the post-meeting work. All recommendations for revisions and additions were then circulated to the full panel for comment and further revision as necessary.

DISCUSSION SUMMARY

· The context set by the moderator was for the panel to imagine that they had been called on to advise a government official who had the legal responsibility to make a hazard determination for a particular exposure, such as whether a substance should be designated a known carcinogenic hazard, or something lesser, such as a possible hazard, which would then be publicized by the government and possibly used for regulatory purposes. For this project, the panel's review of the "London Principles" was organized around the topic of whether the body of studies forming the materials for the conference indicated a causal relationship between induced abortion and breast cancer.

· The points summarized below indicate views expressed by one or more of the discussants; they do not necessarily represent consensus views unless so indicated.

A. Consideration of Biological Information

The Panel's discussion was begun by the two non-epidemiologists with reproductive toxicology experience commenting on the biological basis for judging whether there was a causal relationship at work.

<u>First Discussant</u>

· In the most relevant animal studies, a portion of a group of pregnant rats, of an inbred strain known to be susceptible to 7,12-dimethylbenz(a)anthracene ("DMBA"), a known animal carcinogen, were given hysterectomies while another portion of the group were allowed to go full term and give birth. Subsequently, both portions of the group were administered DMBA. Those that had their pregnancy interrupted responded the same as virgin rats -- *i.e*., the DMBA produced a higher incidence of breast cancer in both virgins and those whose pregnancies had been interrupted. Those that were allowed to

go to full term were protected from breast cancer that might have resulted from the DMBA.

· Studies have also been carried out on human breast development which indicate that pregnancy allows the breast to mature, and therefore the assumption is made that such maturation confers a degree of protection from carcinogens because further cell differentiation, which contributes to carcinogenesis, is limited.

· The assumption has been made that the animal experiments are relevant to the human situation. There are significant flaws in this reasoning. The first is the assumption that the human breast is similarly susceptible to DMBA. The second is that the DMBA was administered after pregnancy rather than before; and if it had been administered before, the reasoning would seem to lead to the conclusion that the cell differentiation that occurs with full-term pregnancy would enhance the carcinogenic effect of the DMBA. In addition, the experiments were carried out only with nulliparous rats.

· To sum up, the reviewer thought that the biological information was significantly incomplete, and that it would not strongly support a causal link with induced abortion, although it could provide some support. The animal data did not either strongly support or strongly detract from a hypothesis that induced abortion increases the risk of breast cancer. The toxicologist also thought that it was important that DMBA would be characterized as an initiator, whereas estrogen would be more likely to be characterized as a promoter. (There was a discussion over the accuracy of this terminology, and one of the epidemiologists noted that in their field they were usually pleased just to know whether an agent acts at an early or later stage, as opposed to understanding the mechanism involved.)

Second Discussant

12

·   The second discussant spoke more to the human biological data. Many reviewers of the issue seem to have assumed that pregnancy is an estrogen-mediated event. In reality, the endocrine biology is far more complex, involving many other hormones, such as progesterone, cortisol, thyroid, and growth hormones. Also, proliferative effects on the breast, including swelling, occur during menstruation. Further complicating the situation, there are psychopharmacologic changes that accompany childbearing, particularly in the hypothalmic-pituitary-adrenal axis, and women modify their behavior in connection with pregnancy -- abstinence from alcohol and tobacco being examples. Still a further consideration is what happens with the immune system during pregnancy. Since the immune system is suppressed during pregnancy, one might suppose that a full-term pregnancy could enhance the opportunity for carcinogenesis. Thus, it seemed naive to regard the biology as supporting a conclusion that induced abortion causes breast cancer based on a focus on estrogen. The changes that occur in early pregnancy, and therefore the effects that could occur from interference with those changes, are not minimal; but pregnancy involves a complex cascade of biological events, and those changes, while significant, are not very different from many other hormonal excursions that occur during other aspects of life.

·   With regard to spontaneous vs. induced abortion, there are some differences in hormone levels in a normal pregnancy vs. one destined to have problems, but they usually do not occur until after seven or eight weeks, whereas most induced abortions occur before that point.

·   It would be difficult to assess whether an abortion during the second trimester would have more impact than one during the first trimester. The biological events that occur as pregnancy progresses involve many factors other than estrogen, such as insulin levels, so that what happens in the second trimester would be different from what happens in the first trimester; but whether it would involve more risk from a biological perspective probably could not be answered.

- Another member of the panel noted that this seemed to mean that it would be very difficult for an epidemiologist to try to figure out what dose-response would mean in terms of this issue.

Panel Discussion

- Both reviewers on the issue of biology agreed that there were substantial gaps in the biological understanding, and that the data did not indicate that there were dose-response effects that would be likely as a result of increasing numbers of induced abortions.

- The reaction of the epidemiologists to the above comments on the relevant biological information was generally that it was useful in the sense that it let them know that there could be biological plausibility, but that the data were equivocal and did not give them good information from which to make decisions on study design, such as how to look at dose-response or timing or which biomarkers were of particular importance. They considered it significant that the biological data did not argue against a possible causal relationship. EMF (electro-magnetic fields) was discussed as an example of an exposure regarding which non-epidemiologists had raised strong arguments against the biological plausibility of a causal relationship.

B.   Comments on Individual Assigned Epidemiologic Studies

1.   Nishiyama et al. (1982, RR of approximately 2.5)

First Reviewer

- Using the existing Principles for this exercise was "torture", because, while much of the content of the Principles was good, the subquestions were filled with compound questions that supposedly required a single Yes or No response.

· The study report was fairly crude by modern standards in this country, and he had to respond to many of the subquestions with a simple "do not know".

· This led the reviewer to comment on the important distinction between study reporting and study quality. This study report did not give sufficient information to answer many questions the reviewer had, but that would not lead him to infer that the study was of poor quality. He would have to contact the investigator(s) and ask about many things. It was noted by the reviewer and others on the panel that some of the most prestigious journals, in which authors/investigators most aim to publish, put strict page limits on articles, so that the description of the study is necessarily truncated. In reviewing a body of literature on an issue, this reviewer would give less weight to a less informative article, not because it was of poorer quality, but because it had a higher degree of uncertainty as published.

· Subquestion A-1(b) was particularly in need of revision. It represented an outdated conception of subject selection. It does not matter whether exposed and unexposed subjects are "comparable at baseline" if the proper adjustments are made. The appropriate question is whether exposed and unexposed subjects are expected to be at comparable risk at baseline with consideration of possible confounding factors.

Second Reviewer

· There were clearly a lot of gaps in the study report, such as not making clear how the control group was comprised, but the reviewer would give some weight to it because there was a large sample size and it considered most of the other risk factors.

· Use of the London Principles showed many of the gaps, but many of the questions in the Principles did not fit well with case-control studies, which this was, and which many of the other studies were.

<u>Panel Discussion</u>

·   All the controls or adjustments in the world will not make things right if the subject selection process is inherently biased.

·   There should be a list of considerations for subject selection that touches on those several areas where big errors are possible.   One must also frame the questions in terms of whether there was a problem with a specific aspect of the study, not simply whether something was considered.

·   It was noted that, from experience in medical practice, women who choose to have a child vs. those who choose to have an abortion have discernibly different behavioral/lifestyle patterns (such as drug use), and this is a significant confounding factor.

·   There was discussion about adjustment for confounding. The issue raised was the validity of many studies stating that they had statistically adjusted for confounding and had found that it made little difference in the outcome, and even giving a very precise figure for the amount of adjustment. One panel member commented that the flaw in this is that the reason adjustment for confounding often does not seem to make much difference is because the investigators have not measured well for key confounders. Another commented that often the search for confounding is like "the drunk looking under the lamp post". Another related this to the current topic by noting that we know there are differences between those who get abortions and those who give birth, but we do not have a handle on those differences as possible sources of bias.

·   There was discussion about the possible need to expand on Principle A-5, which addresses the need to adjust for confounding and bias. Several panelists noted that they thought the Principle was heartening because it made a start on an important issue, but that they thought it

16

should be expanded, for example to say that the investigators should explain how they adjusted.

·   One panelist expressed consternation over the common practice of simply using a standard computer model, such as Mantel-Haenszel, to adjust, characterizing it as "alchemy" and expressing the opinion that what was needed was more sensitivity analysis, because the standard statistical analyses do not begin to touch on the real problems. Sensitivity analysis is what you do when you do not have the data, the measurements. The current issue is a good example of having many possible confounders and biases for which there is no good information that would permit confident adjustments, and therefore the uncertainties would be large, and the sensitivity analysis would reflect those uncertainties.

2.   Howe et al. (1989, overall RR of 1.9)

<u>Both Reviewers</u>

·   The study had some positive aspects, but some substantial weaknesses: It lacked information on most potential confounding factors, and the method of exposure ascertainment was confusing and possibly badly flawed in using fetal death certificates, which would not give good information on spontaneous abortions. As a result, the reviewers would give the study little weight.

·   Another possible weakness was that the study did not seem to have uniform criteria for inclusion/exclusion of cases and controls.

·   Although the Principles were helpful on the better studies, on this one, because the reviewers felt the significant flaws were so obvious, the Principles did not add much. For someone relatively new to the field, the Principles might cause points such as the ones relevant to this

study to be brought to attention; but for an experienced epidemiologist, such significant flaws are conspicuous.

Panel Discussion

- There were comments that the Principles might be trying to do too much in a condensed form. The Principles themselves seemed fine; but it seemed that many of the subquestions did not fit depending on the type of study (cohort or case-control) or the type of exposure, and perhaps what was needed was different layers of principles -- a general set, and then more targeted sets for different types of studies and types of exposures. For example, with regard to the study being reviewed, the subquestion that seemed most pertinent to the significant issues was A-3(g) (on accuracy of exposure measurement or estimation), but it did not capture the issues very well. There seemed to be too much in the subquestions about whether something or other was reported in the study, when the real question was whether there was likely to be bias, such as false negative or false positive attribution of exposure. Another question should be whether the exposure measurement was reflective of the relevant time period. The bottom-line question should be whether the exposure measurements were accurate and relevant.

3. Adami et al. (1990, overall RR of 0.9)

First Reviewer

- That it was a multi-center study was a strength, and it was basically sound methodologically. However, there appeared to be heterogeneity in the case-control sampling and response rates; data collection methods could have influenced results; and there was no stratified analysis by center.

·   The relative homogeneity of the populations in terms of racial makeup and cultural factors seemed to be a distinct advantage for this study, as opposed to one conducted in the United States, where one would expect large differences among the study populations with regard to such factors. The Principles did not seem to capture this point well.

·   The main Principles seemed to be generally sound, but, again, the reviewer had difficulty with the subquestions, in large part because they often contained compound questions, so that it was difficult to give a single Yes or No answer.

·   For a number of Principles subquestions, such as blinding and quality control, one could not answer the question, but one would assume that that aspect of the study was done right if they knew the study team was reputable. The quality control aspects are important, but no journal is going to publish that much detail. There is a question of what is realistic to expect, particularly if it is a routine matter, as opposed to, say, a formal validation study. One would like to know there is a more detailed write-up somewhere that documents everything, but one cannot expect such detail in the journal.

Second Reviewer

·   Under Principle A-6, the reviewer thought that subquestions (e) (explaining contradictory or implausible results) and (f) (exploring alternative explanations) were largely irrelevant, because an experienced reviewer could do this if the data from the study were reported correctly. It could earn the authors a pat on the back, but really what was important was the particular findings of their study. If they do what is suggested in the Principles, it does not change the quality of their data. However, it could help another reviewer in thinking about a particular aspect of their study and how much weight to give the study.

·   Under the same Principle, subquestions (g) and (h) (interpretation of results and discussing public health implications) should actually be discouraged in the reporting of individual studies, whereas the Principles seemed to be encouraging it. The presence of this type of discussion in an individual study would raise questions regarding the motives and biases of the study author(s).

Panel Discussion

·   This seemed to be the only study they had ever seen which did not find a protective effect with pregnancy, which raised concerns.   In addition, one could not tell from the article what the latency was. Also, there were concerns about the range of the confidence intervals, and it seemed that the study was not designed to assess the issue under consideration.

·   It would be gratifying to have the opportunity to really refine the Principles through several iterations.

4.   La Vecchia et al. (1993, RR of 0.9)

Both Reviewers

·   There were problems with selection of controls because it was a hospital-based study with controls drawn out of other patients in the hospital. The use of cases of all ages without looking at age-specific risks probably also biased results.

·   Additionally, the study, like the previous one, was not really designed to address the issue under consideration.

·   In general, the study was not very useful, and the London Principles were simply in line with how an experienced epidemiologist would analyze it.

<u>Panel Discussion</u>

· There is some schizophrenia inherent in the Principles at present. On the one hand, they appear to be informative to non-epidemiologists; but it would be very dangerous to give non-epidemiologists the idea that they could apply the Principles and make accurate appraisals. On the other hand, the Principles as they stand, while they should be applied by expert epidemiologists, seem to largely reflect basic precepts that all expert epidemiologists should be familiar with. If the Principles are to further the state of the art, or bring more consistency to individual epidemiologic studies, evaluation of studies, and evaluation of bodies of studies, they should be refined. It was emphasized that the Principles should not be employed by non-epidemiologists in an official capacity. This would be like trying to have epidemiologists evaluate toxicology studies. Assuming the Principles are to be used by epidemiologists to advise non-epidemiologists, the aim should be to make evaluations more systematic and consistent; but the Principles should be advisory, not dogmatic.

· Doubt was expressed about the statement in Principle A-6 that a study should be of sufficiently high quality to be publishable in a peer-reviewed journal. Publication is not much of a quality control these days; and much that is not published may be of higher quality than that which is published.

· Epidemiologists like to publish their best work in journals that will give them sufficient space to provide the details of their analysis, which are usually the more specialized journals rather than the ones intended for a wider audience.

· The Panel seemed to agree that what was really needed was three separate sets of principles or questions: One consisting of principles or questions for epidemiologists to use in evaluating individual studies;

21

the second for epidemiolgists evaluating a body of studies; and a third for lead risk assessors and risk managers to employ in consultation with the epidemiolgists and other risk assessors.

5.   Laing et al. (1993, overall RR of 3.1)

First Reviewer

·   The use of histologically confirmed cases and large study size were strengths.

·   But there were what could be called fatal flaws that made it of minimal or no utility for risk assessment purposes. The controls were selected in a way that put them at the extreme of non-comparability; and the questioning strategy opened up the likelihood of response bias and bias due to confounding from use of oral contraceptives (cases were probably asked about oral contraceptive use, while the controls probably were not)-- in other words, problems with exposure ascertainment.

·   With regard to the London Principles, they added little that was not already apparent to an experienced epidemiologist on first reading.

Second Reviewer

·   Agreed with the first reviewer's points.

Panel Discussion

·   The London Principles do not allow for gradations of quality; everything is just yes, no, or don't know/not applicable. There should be a way to allow for at least a scale of three of four points on a particular aspect of study quality, such as control of confounding,

exposure measurement, and suitability of the control group. This is not the same as giving an overall quality score for a study; it would allow for integration across diverse dimensions. This would be for the purpose of assisting the epidemiologist in evaluating the individual studies, and then he/she would make a judgment as to utility for risk assessment based on the issue-by-issue evaluation of the study.

6.   Daling et al. (1994, overall RR of 1.36)

First Reviewer

·   In general, this appeared to be a careful study; however, there were some minor weaknesses or uncertainties.

·   It was not clear from the study report how soon after diagnosis the cases were interviewed. It appeared that there might have been a substantial time lapse, so that they would have had more time to speculate on the cause of their disease, which could have introduced more potential for recall bias.

·   It would have been useful to have information on risk factors for non-respondents vs. respondents.

Second Reviewer

·   The efforts to obtain self-reported data on induced abortion seemed to be as good as it could get -- the investigators were up against the limits of what could be done.

·   It was a good example of the state-of-the art in case-control studies; but nevertheless the mundane, familiar, annoying potential sources of error remained. Non-response is not likely to be random, and the reasons may well not be the same for cases as for controls. But there does not seem to be a way to fix that. There is also still the potential for lack of truthfulness or inaccurate recall in response to questions

about having had an induced abortion, and that also appears unresolvable.

Panel Discussion

· There could be an issue of statistical bias in this study of a type that is not generally recognized: Adjustment for multiple confounding factors in a small subgroup can cause the risk ratio to inflate. This artifact needs to be better understood by investigators.

7.   Lipworth et al. (1995, overall RR of 1.5)

Both Reviewers

· It was difficult to tell how much weight to give this study. One reviewer would give it a fair amount of weight; the other would be faced with a lot of uncertainty. The analysis in the study report seemed to be fairly good; but it was difficult to tell how the controls were selected, and it appeared that the method of selection was peculiar. The controls did not seem to be well-related to the population from which the cases came. This goes to the suitability of controls and the possibility of selection bias. And yet, the way the study was presented, and the fact that the usual risk factors seemed concordant between the cases and controls, gave one the feeling that perhaps the study was better than it was entitled to be regarded. Control selection is probably the most important aspect of a case-control study; and if it is flawed, the whole study will be flawed.

· In addition, the study was not designed to examine the issue under consideration.

· The Principles tended to confirm the problem with controls that was noted at the outset. If one had to genuinely confront and respond to the Principles, it could force one to be more rational about the control selection issue and give the study less weight than one might

24

otherwise be inclined to give it after forming a subjective impression from a simple reading.

Panel Discussion

·   On the matter of control selection, this again is an area where the Principles need improvement. The way it is expressed is outmoded and flawed. The issue is not whether the cases and controls are comparable; it is whether they were selected from the same population as the cases in some identifiable fashion, or by random method.

8.   Daling et al (1996, RR in range of 1.2 to 2.0)

Both Reviewers

·   A strength was the use of multiple centers to increase sample size and the heterogeneity of the population.

·   The main problem with the study was that the control sampling method resulted in an unusual number of poor matches -- differences in matching cases and controls across the different geographic sites -- but the investigators recognized this issue and appeared to have made a valid statistical adjustment. Control selection was by random digit dialing, and it appeared that making the calls cross-country could have been a problem.

·   The study was well-conducted and would be entitled to equal weight with other well-conducted retrospective studies.

Panel Discussion

·   It was meritorious that the problem with controls was noted in the article.

·   But it is hard to correct for a control selection problem and be assured of having eliminated bias. It is much better to do it right in the first place.

·   This is a good example of a study that could get a reduced quality score because it identified a problem, whereas another report might actually be of lower quality but perceived as of higher quality because the investigators did not flag a problem and, using the Principles, one would mark down "Not Known".

·   One advantage in using the Principles is that the "Not Known"s could lead one to go back and check with the investigator(s), which is something that should be encouraged, particularly when one is doing a meta-analysis.

9.   **Lindefors-Harris et al. (1991, response bias study comparing two studies with range of overall RR's of 1.1 to 2.0)**

First Reviewer

·   The reviewer had recently done a study on fat intake and breast cancer in which the controls apparently did not under-report fat intake.

·   The subject study, which attempted to look for response bias, appeared to be fairly solid because it compared responses with abortion registry information.

·   However, the investigators calculated a relatively crude response rate, and they did not attempt to consider potential confounders.

Second Reviewer

·   A possible weakness was treating false positives similarly to false negatives, when the psychological factors appear to be dissimilar -- in other words, it would be odd for someone to report that they had an induced abortion if they had not.

Panel Discussion

·   Much of the response difference found in the study came from cases that reported an abortion that did not show up in the registry.   There was also a comment that it had been heard that the reporting to this registry was poor.

·   Age is an important factor in recall, of course. Many older women (post-menopausal) cannot remember that they had a spontaneous abortion that is recorded in their medical records. In the subject study, women were being asked to recall quite a long time ago and remember fairly accurately when an induced abortion occurred.

·   In the case of teratogenesis studies, the discussants thought that the literature showed that both cases and controls did a poor job of recalling, and it was mostly non-differential, although the cases did slightly less under-reporting.   There was little in the way of false positives. There is both random bias and differential bias involved; but the bias seems to be towards the null.

·   But the subject studies introduced unusual factors such as religion and social desirability, and those factors seem to depend on the cultural milieu, so that there is a question over the generalizability of the results from one culture or region to another.

·   Perhaps the only way to handle the issue of potential response/recall bias as a general proposition is to do sensitivity analysis.

        10.   Rookus et al. (1996 study in which authors concluded that regional reporting bias could account for an RR of 1.9)

·   The study was not entitled to much confidence. The power of the study was very low; the response rate was low; and the comparison between reporting of induced abortions and reporting of duration of oral contraceptive use (rather than any use of contraceptives) was dubious, because women seem to have a difficult time recalling the latter accurately.   If they were going to mis-report on oral contraceptive use, one would expect that they would deny any use, rather than distort the duration of use.

Second Reviewer

·   Agreed with the above observations regarding weaknesses, but thought the concept of the study -- comparing responses from populations with different religions or cultures -- was interesting and promising.

11.   Melbye et al. (1997 registry study with overall RR of 1.0, and RR of 1.39 at >12 wks.)

First Reviewer

·   There was a great potential for exposure misclassification that would tend to bias towards the null. The study used subjects of all ages, and the incidence of breast cancer increases with age, and many of the older women would have had induced abortions 25 or 30 years before, but the data only went back to 1974, and they assumed that there were no induced abortions for subjects before then. So there was bias in the data, even though it purported to be unbiased because it was registry data.

· It was not clear how the crude risk of 1.4 had been adjusted down to 1.0.

· The study included nulliparous women without analyzing them separately, and there is a lot of evidence that such women are at higher risk. It also included nulligravid women, who would be likely to lack comparability to gravid women for a number of reasons. It would have been useful to see more subgroup analysis.

Second Reviewer

· Agreed with the main points made by the first reviewer, and noted that this study appeared to be like a mirror image of the other studies they had reviewed, in that it had none of the weaknesses in common and none of the strengths in common.

Panel Discussion

· The main issue here was a truncated exposure history. The London Principles contain a lot of questions about exposure under Principle A-3, but none appear to address the issue of exposure data that are incomplete in a temporal sense. There should be a question about whether there were substantial missing data, and, if so, how that was handled. This is often a critical issue. There are textbooks on how to handle missing data; but some investigators do bizarre things when faced with missing data.

· Some interested parties (outside the panel) had apparently focused on this study as settling the response bias issue; but at the same time it created other problems. So in epidemiology things are often not as simple as they seem.

12. Brind et al. (1996 meta-analysis with synthetic RR of 1.3)

First Reviewer

· Statistical principles underlying a sound meta-analysis are no different from those for an individual study. Many reviewers conducting meta-analyses have ignored standard statistical principles and invented, or utilized, statistical approaches which would not withstand criticism if used in an individual study.

· Meta-analysis is not about combining evidence to arrive at a synthesis -- *i.e.*, a single highly precise number. One should expect a good meta-analysis to examine for disparities, contradictions, etc.

· Selecting for a meta-analysis the "best" studies based on apparent quality is a distortion -- it is like deciding to select study subjects based on whether you like their appearance. The reviewer presented a quotation from Kelsey et al.: "If the data show a marked heterogeneity, then attempts to summarize on a single rate ratio can obscure important features of the underlying rates and should be avoided."

· Heterogeneity means differences in study results beyond what would be expected by chance, whether they are case-control or cohort. When you see heterogeneity in the results, you have to go back and look for heterogeneity in the study design and methodology and attempt to discover why it is present.

· The subject meta-analysis did not attempt to look for heterogeneity in the results in a meaningful way. Meta-analysis must be an analysis. There was almost no statistical analysis in this paper, and the reviewer would rate its quality as very poor.

· An initial serious flaw was that the synthetic analysis, when the data were re-analyzed and replicated, showed that the authors utilized a

fixed effects summary rather than a random effects summary. The study concluded that the included studies showed consistent significant positive associations, but it did not examine the statistical agreement. For example, it did not analyze for differences in confounding.

· The results of the meta-analysis look precise, but if just one subsequent study is added -- the Melbye et al. study (see above) -- the synthetic number drops from 1.3 to 1.1. In other words, the results, even as they are, are extremely fragile.

· If an analysis for global heterogeneity is done (*i.e.*, assuming the studies are all perfect and just using their results), the range of heterogeneity is extreme -- that is, the analysis indicates that the results are very likely to be due to chance. Although the mean is about the same as the study's synthetic number, the distribution from the random effects analysis is extremely broad -- from 0.67 to 2.9. The 95 percent confidence interval in the random effects distribution is usually misinterpreted. What it means is that it is only a confidence interval for the mean of the random effects distribution, and that is of virtually no scientific interest.

· The bottom line was that there was so much heterogeneity apparent from just going no further than taking individual study results at face value that a summary number was not justified. The meta-analysis was fatally flawed.

· A summary chart showing the confidence intervals of the various studies is a necessary starting point, but it is only a starting point.

· If one were to examine the various studies for heterogeneity just by geographical location, the amount of heterogeneity is extremely large and unaccounted for and cannot be due to chance.

·   The authors did some sub-analysis and stratification, but they completely misinterpreted it. They found that the results changed little in the sub-analyses, and from that they concluded that the sub-analyses supported a finding of a real effect. Just the opposite should be the case. If there is great heterogeneity in results to start off with, the fact that the sub-analyses show little variation means that they do not account for the great initial heterogeneity. The homogeneity in the sub-analyses implies that there is tremendous, significant, unexplained heterogeneity. Additionally, if the biological hypothesis is correct, one would expect significant heterogeneity in the sub-analyses, but it is not present.

·   It is probably less deleterious simply to do a narrative analysis rather than a synthetic meta-analysis when there is substantial initial heterogeneity in study results, unless one is prepared to do a great deal of hard work in analyzing the data to see if they can explain the heterogeneity. Using a random effects model to summarize study results is not as misleading as using a fixed effects model, but it still sidesteps the main issues. One of the problems with using a random effects model is that it downweights the larger studies and is more sensitive to the smaller ones.

·   The problem with unpublished studies (sometimes referred to as the file drawer problem) is very real. Usually, if you are doing a meta-analysis and you beat the bushes you will come up with unpublished work. It usually tends to be null, but it could go either way.

Second Reviewer

·   Had thought the analysis was objective and unbiased and looked reasonable, particularly with regard to inclusion of all the relevant studies. Also had been under the impression that a random effects analysis had been used.

·   Although the summary statistical synthesis was followed by a qualitative review, the qualitative review was really only helpful to the extent it gave clues as to what should be looked at in a statistical analysis of heterogeneity. The qualitative analysis was not systematic because it did not look at the same factors across all the studies.

·   Study attributes should be organized systematically and then compared/contrasted and analyzed. It should be done like a single study, in which the individual studies are like individual subjects in the single study.   You analyze for what are the important likely confounders or effect modifiers, possible biases, and you do subgroup analysis and stratification.

Panel Discussion

·   It should be taught that meta-analysis is a comparison, not a combination, of results of different studies -- that is, detection and description of systematic variation among study results.

·   If the studies are all fairly homogenous with tight confidence intervals, then you would not even bother with meta-analysis; on the other hand, it should be kept in mind that the consistency could be due to a consistent bias.

·   One panel member brought a copy of a 1997 article by Wingo et al. on abortion and breast cancer that was not originally included in the workshop materials. It was characterized as a systematic narrative review.   It considered the body of literature to be too inconsistent to draw conclusions from, and it suggested the future studies focus on certain identified issues. It was commented that this identification of gaps that should be addressed in future studies was particularly useful and should be a function of review articles.

33

·   It was noted that it is surprising on a controversial issue to see reviewers come to such firm conclusions based on such a weak (synthesized) relative risk, in this case 1.29, when some reviewers will take a weak relative risk as a reassurance that there is no risk.

C.   Discussion of Overall Views and Desired Work Product

·   It is important that the majority of the studies were not designed to address the issue.

·   Some of the epidemiologists commented that some aspects of Hill's factors are considered outdated or faulty, such as endpoint specificity and dose-response.   The toxicologist on the panel disagreed, contending that dose-response and specificity should be considered very important, if not essential, factors in considering causality.[10]

---

[10]   There were further comments exchanged on this subject following the Denver meeting; however, it is not possible to say whether the issues were resolved completely. Some of the epidemiologists clearly attach less significance to some of Hill's factors, such as dose response and specificity, than this toxicologist. Several comments by the moderator seem in order here: First, some of the epidemiologists seemed to view "Hill's factors" (they dislike the term "criteria" as connoting hard and fast rules) as referring to dose-response as a simple monotonic upward gradient; whereas the toxicologist, and probably most or all toxicologists, are of the view that this factor pertains to some recognizable <u>pattern</u> of dose-response, which might in different cases be, as examples, a "J", "hockey-stick", threshold, or saturation pattern -- but in any event not a random or zig-zag response.   Second, some of the epidemiologists seemed to view Hill's factor of specificity as pertaining to a single organ or disease endpoint; whereas the toxicologist is again looking for a recognizable and plausible consistency, which might involve multiple sites with one or more sites predominating -- but again with a pattern rather than a random response. (It seemed to be agreed that exposures to heterogeneous mixtures are a different matter.) There appears to be considerable overlap among "Hill's factors" of dose-response, specificity, consistency, coherence, and biological plausibility, which can lead to difficulties in discussing individual factors. It also appears that there is a need for epidemiologists and toxicologists/pharmacologists to communicate better on these issues, and that the London panel's recommendations to convene multi-disciplinary panels to evaluate the evidence was well-advised. A somewhat detailed discussion of Hill's factors appears in Appendix B of the London report; and there is less specific reference to some of

·   Done well and adhered to, the Principles could make IARC-type hazard determinations more transparent than simply the result of a majority show of hands by a group of experts.

·   There is a large body of often-overlooked literature, which has been successfully utilized, for rationally integrating expert views, such as those arising from a meta-analysis. Reference was made (non-specific) to a classic example of an issue over whether a drug was causing an outbreak of disease, and various experts expressed their views, and when those views were rationally analyzed using such decision principles, they found there was much more uncertainty than they had previously realized.

·   As epidemiologists, they would be comfortable in converting the existing hazard Principles into two separate sets: One for evaluation of individual studies; and one for evaluation of a body of studies. Then there could be a third set of principles or questions designed to help integrate the first two sets of information for purposes of making a policy decision.

·   Evaluating a body of studies goes beyond the formal training of many epidemiologists. Their training focuses mainly on conducting and evaluating individual studies.

·   Much of what has been written on establishing a causal relationship or statistically analyzing a body of studies is flawed. Some of the people expounding on this subject are not trained in the appropriate body of literature and methods. Most epidemiologists should undergo further training on how to conduct a proper meta-analysis for the purpose of examining whether there is a causal relationship.

---

those factors in Appendices D, E, G, and H.

· There is a body of decision analysis literature that addresses how one maintains objectivity and quantifies uncertainty. There are experts in this type of analysis who are neither epidemiologists nor statisticians. The job of such decision analysts is to draw out the specialists, the epidemiologists in this case, without the specialists necessarily knowing what they are up to, and what principles or methods they are utilizing. To draw up a set of principles for utilizing such a process would probably not be an "off the shelf" effort, but would require convening a mixed group of epidemiologists and other experts who had this particular type of integrative decision logic expertise.

· There was some disagreement over whether epidemiologic expertise should be regarded as ending with the analysis of individual studies.

· The broad goal of the Principles should be to assist in avoiding significant mistakes, not dictating exactly how to do analysis.

· The London meeting and Principles and report were a good first step; this panel was now willing to try to refine that work; and then a larger group should conduct a further review.

· The panel (the four epidemiologists remaining at the very end of the meeting) agreed to allocate assignments so that two would divide up work on draft revisions of the six principles and subquestions for evaluating individual studies, and the other two would initially work on the meta-analysis principle (B-6). The non-epidemiologists and Federal Focus staff would draft a third set of principles or questions aimed at decision integration for hazard identification and risk assessment. The drafts would then be circulated for review by the full panel.

36

RECOMMENDED REVISIONS TO
THE "LONDON PRINCIPLES"

The following suggested revisions to the London Principles were developed by the Denver panel participants and agreed upon except where exceptions or alternative suggested wording are noted. London Principles text recommended to be deleted is crossed through, and recommended new or revised text is in bold italics.

~~A.~~ *I.* PRINCIPLES *AND QUESTIONS* FOR EVALUATING
AN EPIDEMIOLOGIC REPORT FOR
CAUSE-EFFECT RELATIONSHIP[11]

Revisions to Principles A-1 to A-3

Principle I-1.   The population studied should be pertinent to the risk assessment at hand, and it should be representative of a well-defined underlying cohort or population at risk.

a.   Were study subjects representative of exposed and unexposed persons (cohort study), or of diseased ~~and non-diseased persons~~ *and*

---

[11] Several of the panelists wished to note that these principles and questions, and the ones for meta-analysis, could be used to evaluate whether there was a causal relationship with a beneficial, as well as an adverse, health outcome attributable to exposure.

of persons at risk of getting disease[12]
(case-control study)? ***If not, were data***
collected on major risk factors for
the disease or condition under study
to allow appropriate adjustment in
the analysis for potential confounders
and evaluation of effect modifiers   Yes   No   NK   NA

---

[12] Two panelists suggested adding following "disease":   "who are sampled from the source population of the cases".

in the analysis? [13]   □   □   □   □

b. ~~To minimize bias~~ Were exposed and
unexposed persons comparable "at
baseline" ~~(cohort study), or were
cases similar to controls,~~ prior to
exposure, with respect to major risk
factors for the disease or condition
under study? *If not, were data collected
on major risk factors for the disease or
condition under study to allow
appropriate adjustment in the analysis
for potential confounders and
evaluation of effect modifiers in
the analysis*?   □   □   □   □

Principle I-2. Study procedures should be described in sufficient detail, or
available from the study's written protocol, to determine whether appropriate
methods were used in the design and conduct of the investigation.

a. ~~To minimize the potential for bias,~~
Were interviewers and data collectors
blind to the *exposure status and*
case/control status of study subjects
and to the hypothesis being tested?   □   □   □   □

b. Were there procedures for quality
control in place for all major aspects
of the study's design and
implementation (*e.g.*, ascertainment

---

[13] One panelist suggested that "for potential confounders . . . ." be replaced
with "for selection bias. . . ."

and selection of subjects for study,
methods of data collection and
analysis, follow-up, etc).

	Yes	No	NK	NA
	☐	☐	☐	☐

c. Were the effects of nonparticipation,
~~a low response rate~~ **nonresponse**,
or loss to follow-up taken into account in
producing the study results? ☐ ☐ ☐ ☐

Principle I-3. The measures of exposure(s) or exposure surrogates should be: (a) conceptually relevant to the risk assessment being conducted; (b) based on principles that are biologically sound in light of present knowledge; and (c) properly quantitated to assess dose-response relationships.

a. Were well-documented procedures
for quality assurance and quality
control followed in exposure
measurement and assessment (*e.g.*
calibrating instruments, repeat
measurements, re-interviews, tape
recordings of interviews, etc.) ☐ ☐ ☐ ☐

b. Were measures of exposure
consistent with current biological
understanding of dose (*e.g.*, with
respect to averaging time, dose rate,
peak dose, absorption via different
exposure routes)? ☐ ☐ ☐ ☐

c. If there is uncertainty about
appropriate exposure measures, was
a variety of measures used (*e.g.*,
duration of exposure, intensity of
exposure, latency)? ☐ ☐ ☐ ☐

d. If surrogate respondents were the
    source of information about exposure,

    was the proportion of the data they
    provided given, and were their
    relationships to the index subjects
    described? ***Were risk estimates
    excluding surrogate respondents***   Yes   No   NK   NA
    presented?   ☐   ☐   ☐   ☐

e. To improve study power and enhance
    the generalizability of findings,
    was there sufficient variation in the
    exposure among subjects?   ☐   ☐   ☐   ☐

f. Were ~~correlated~~ ***other*** exposures
    measured and evaluated to assess the
    possibility of competing causes,
    confounding, and potentiating effects
    (synergy)?   ☐   ☐   ☐   ☐

g. Were exposures measured directly
    rather than estimated? If estimated,
    have the systematic and random
    errors been characterized, either
    in the study at hand or by reference
    to the literature?   ☐   ☐   ☐   ☐

h. Were measurements of exposure or
    human biochemical samples of
    exposure made? ~~Was there a
    distinction made between exposures
    estimated by emission as opposed
    to body absorption?~~ ***Was there a
    distinction made between external***

41

exposure and internal dose, if
appropriate?   ☐   ☐   ☐   ☐

i.   If exposure was estimated by
     questionnaire, interview, or
     existing records, was reporting bias
     considered, and was it unlikely to   Yes   No   NK   NA
     have affected the study outcome?   ☐   ☐   ☐   ☐
j.   Was there an explanation/understanding
     of why exposure occurred, the
     context of its occurrence, and the
     time period of exposure?   ☐   ☐   ☐   ☐

Revisions to Principles A-4 to A-6

~~Principle~~ *I-4.* ~~Study outcomes (endpoints) should be clearly defined, properly
measured, and ascertained in an unbiased manner.~~ *Definition and
measurement of study outcomes:*

a.   ~~Was the outcome variable a disease
     entity or pathological finding rather
     than a symptom or a physiological
     parameter~~? *Was the operational
     definition of the health outcome
     provided?[14]*   ☐   ☐   ☐   ☐

b.   ~~Was variability in the possible
     outcomes understood and taken into
     account -- *e.g.,* various
     manifestations of a disease
     considering its natural history?~~

---

[14] Two panelists were unsure whether this covered sufficiently whether the
health outcome measured was the most relevant one.

42

Was the health outcome defined using objective rather than subjective criteria? □ □ □ □

c. ~~Was the method of recording the outcome variable(s) reliable — *e.g.*, if the outcome was disease, did the design of the study provide for recording of the full spectrum of disease, such as early and advanced stage cancer; was a standardized classification system, such as the International Classification of Diseases, followed; were the data from a primary or a secondary source?~~ *Was the health outcome assessment reliable?*

Yes	No	NK	NA
□	□	□	□

d. ~~Has misclassification of the outcome(s) been minimized in the design and execution of the study? Has there been a review of all diagnoses by qualified medical personnel, and if so, were they blinded to study exposure?~~ *Was the health outcome assessment valid?* □ □ □ □

e. *Were the component information required for the health outcome assessment valid?*[15] □ □ □ □

---

[15] Two of the panelists wished to note in connection with this subquestion, and also (c) and (d), that if more than one surrogate were used to assess a given

f. *Were the evaluation criteria
    for the health outcome applied
    equally to the exposed and*   Yes   No   NK   NA
    unexposed individuals?   ☐   ☐   ☐   ☐

Principle I-5. ~~The analysis of the study's data should provide both point and interval estimates of the exposure's effect, including adjustment for confounding, assessment of interaction (*e.g*, effect of multiple exposures or differential susceptibility), and an evaluation of the possible influence of study bias.~~ ***Data analysis methods and presentation:***

a. ~~Was there a well-formulated and
    well-documented plan of analysis?
    If so, was it followed?~~ ***Were the
    analytic methods clearly described?***   ☐   ☐   ☐   ☐

b. ~~Were the methods of analysis
    appropriate? If not, is it reasonable
    to believe that better methods would
    not have led to substantially
    different results?~~ ***Were the data
    analysis methods appropriate to
    address the study question*?**   ☐   ☐   ☐   ☐

c. ~~Were proper analytic approaches,
    such as stratification and regression
    adjustment, used to account for
    well-known major risk factors
    (potential confounders such as age,~~

outcome, one should ask whether the surrogates gave roughly the same answer both in terms of direction and effect size.

~~race, smoking, socio-economic~~
~~status) for the disease under study?~~
If necessary, was confounding
controlled using appropriate
statistical methods? □ □ □ □

d. ~~Has a sensitivity analysis been~~
~~performed in which quantitative~~
~~adjustment was made for the~~
~~effect of unmeasured potential~~
~~confounders,~~ *e.g.*, ~~any unmeasured,~~
~~well-established risk factor(s) for~~
~~the disease under study?~~ *Was*
completeness of control for confounding
addressed through sensitivity analysis Yes No NK NA
or other analytic approaches? □ □ □ □

e. ~~Did the report avoid selective~~
~~reporting of results or inappropriate~~
~~use of methods to achieve a stated~~
~~or implicit objective? For example,~~
~~are both significant and~~
~~non-significant results reported in a~~
~~balanced fashion?~~ *Was sufficient*
consideration given to effect modifi-
cation, given the study question? □ □ □ □

f. ~~Were confidence intervals provided~~
~~in the main and subsidiary analyses?~~

f. *Were point and interval measures[16] of*
effect provided? □ □ □ □

_____

[16] One panelists suggested changing "measures" to "estimates".

Principle A-6. ~~The reporting of the study should clearly identify both its strengths and limitations, and the interpretation of its findings should reflect not only an honest consideration of those factors, but also its relationship to the current state of knowledge in the area. The overall study quality should be sufficiently high that it would be judged publishable in a peer-reviewed scientific journal.~~ *Discussion and interpretation of study results:*

a. ~~Were the major results directly related to the a priori hypothesis under investigation?~~   Yes   No   NK   NA   ☐   ☐   ☐   ☐

b. ~~Were the strengths and limitations of the study design, execution, and the resulting data adequately discussed?~~   ☐   ☐   ☐   ☐

c. ~~Is loss to follow-up and non-response documented? Was it minimal? Has any major loss to follow-up or migration out of study been taken into account?~~   ☐   ☐   ☐   ☐

d. ~~Did the study's design and analysis account for competing causes of mortality or morbidity which might influence its findings?~~   Yes   No   NK   NA   ☐   ☐   ☐   ☐

e. ~~Were contradictory or implausible results satisfactorily explained?~~   ☐   ☐   ☐   ☐

f. ~~Were alternative explanations for the results seriously explored and discussed?~~   ☐   ☐   ☐   ☐

g. ~~Were the Bradford Hill criteria (see Appendix B) for judging the~~

~~plausibility of causation (strength of association, consistency within and across studies, dose response, biological plausibility, and temporality) applied when interpreting the results?~~ ☐ ☐ ☐ ☐

~~h. What are the public health implications of the results? For example, are estimates of absolute risk given, and is the size of the population at risk discussed?~~ ☐ ☐ ☐ ☐

		Yes	No	NK	NA
a.	*Are there major threats to the validity of the study?*	☐	☐	☐	☐
b.	*Are the exposed and unexposed groups comparable in baseline (pre-exposure) risk of disease, or have any differences been accounted for in the analysis?*	☐	☐	☐	☐
c.	*[For case-control studies only] Are the controls likely to be representative of the source population from which the cases arose, or have any differences been accounted for in the analysis?*	☐	☐	☐	☐
d.	*Is non-response or loss to follow-up likely to have introduced substantial bias?*	☐	☐	☐	☐
e.	*Are missing data likely to have introduced substantial bias?*	☐	☐	☐	☐
f.	*Is exposure measurement error likely*				

to have introduced substantial bias?	☐	☐	☐	☐

g. ***Is health measurement error likely to have introduced substantial bias***? ☐ ☐ ☐ ☐

h. ***Is error in measuring confounders likely to have introduced substantial bias***? ☐ ☐ ☐ ☐

i. ***Is random error likely to have produced substantial inaccuracy in the measure of effect***?

	Yes	No	NK	NA
	☐	☐	☐	☐

II. PRINCIPLE FOR EVALUATING A BODY OF STUDIES ("META-ANALYSIS")

~~**Principle B-6. A properly conducted meta-analysis, or preferably an analysis based on the raw data in the original studies, may be used in hazard identification and dose-response evaluation when such combination includes an evaluation of individual studies and an assessment of heterogeneity. The combined results ought to provide, more than any single study, precise risk estimates over a wider range of doses. Before using these tools, the gains should be judged sufficient to justify potential errors in**~~
~~**inference resulting from combining studies of dissimilar design and quality.**~~

~~**Discussion of Principle B-6   This is one of the more important principles because meta-analysis of multiple epidemiologic studies has become more common. Although the principle refers to "properly conducted meta-analysis", there are currently no generally accepted standards for employing it (nor is it mentioned in the 1986 EPA risk assessment guidelines that are currently in effect 17), although when used quantitatively it involves application of standard statistical methods. This principle does state several minimal standards for employing meta-analysis, however. Even applying these minimal standards, some felt that meta-analysis was still fraught with potential for error, and it was remarked that if meta-analysis (based on studies as they are often currently conducted) were used as the final determinant in risk assessment, "we are in trouble". Meta-analysis~~ requires consideration of the quality of the various ~~studies, but it is very difficult and controversial to attach weights to studies for~~

quality. Ranking for quality is not a black and white exercise; it requires a lot of expert judgment by epidemiologists.

Epidemiologists feel more comfortable in combining results from randomized clinical trials, when they follow similar protocols; but for observational studies, heterogeneity can create a "morass". If in the future epidemiological studies were to become more like randomizedclinical trials in the transparency and consistency of their protocols (e.g., what the dose metric will be, how exposure and outcome will be measured), they would become more suitable fora well-conducted meta-analysis.

Principle B-6 incorporates three minimal standards. The first, which applies to both hazard identification and dose-response assessment, is that the reviewer should examine the individual studies carefully for quality and for explanations for study differences rather than combining study results simplistically to get a single point estimate. Within this sub-principle there is a stated preference for analysis by "pooling" of raw data where possible. The second is that, for dose-response assessment, combining thedata should produce gains in terms of better risk measures with narrower confidence intervals over a wider range of doses. The third is that one should be mindful of the hazards of meta-analysis, and take into consideration that the body ofstudies for a particular agent may contain so many unexplainable inconsistencies as a consequenceof differences in quality, design, and results that application of a meta-analysis methodology may be counter-productive and may mislead by obscuring significant uncertainties. If there is one very good study, nothing may be gained by combining it with others; in reality, combining may makethe evaluation weaker. All of this requires agreat deal of expert judgment.

"Meta-analysis" as it is performed currentlyoften means summarizing relative risks weightedaccording to sample size, while an analysis that combines the raw data extracted from individual studies as if one were conducting a single large epidemiologic study is often referred to as"pooling". Principle B-6 indicates a preferencefor "pooling" because it allows for a better evaluation of possible systematic flaws or weaknesses, whereas meta-analysis can amplify flaws or weaknesses (for example, by reinforcing biases in different studies); however, it cannot always be done because the raw data may not be available from all of the studies, and it is difficult to apply for dose-response purposes because many studies will not have sufficient exposure detail. Examples given of situations where there was sufficiently good exposure data, in terms of detail and uniformity, to

49

allow pooling were the British analysis of EMF and childhood leukemia, where all the studies used the same method of measuring exposure, and the risk assessment for airborne radon, where the eleven uranium mining studies quantified exposures in a similar way. It is important that the combined analysis cover all studies, whether positive, negative, or inconclusive.

One participant raised the question of whether, after applying meta-analysis or pooling and finding a weak relative risk, the result would be considered stronger than a similar level of relative risk estimated from a single good study. This was considered a difficult question, and some felt that the weak relative risk indicated by the meta-analysis would not be considered more significant unless there was some confirmatory animal data regarding the shape of the dose-response curve and biological plausibility, and even then they would consider the results "equivocal".

__Principle II__: Meta-analysis is a method for combining and contrasting results from different studies that can be used as an aid in the review of evidence in a particular area. A properly conducted meta-analysis, or when possible, an analysis based on the raw data of original studies, may be used in evaluating the epidemiologic evidence concerning the exposure-disease relation under investigation. Such a combination of evidence across epidemiologic studies should follow a rigorous protocol that includes a thorough evaluation of the methods and biases of individual studies, an assessment of the degree of conflict among studies, development of explanations for any conflicts, and, if the evidence is consistent, a quantitative summary of that evidence. The outcome of a meta-analysis of epidemiologic studies can provide more precise assessment of the inter-study variation in the association under investigation, as well as the reasons for the differences observed across individual studies. When the evidence is conflicting, the primary goal of a meta-analysis should be identifying and estimating differences among study-specific results, rather than estimating an overall quantitative summary of the association across studies. With this goal, problems in combining evidence from studies that are dissimilar in design and quality can be minimized.

Discussion of Principle II

This is an important principle for epidemiology in regulatory risk assessment, particularly because meta-analyses are commonly used in policy formulation. The principle begins with a short definition of meta-analysis and then refers to a "properly conducted meta-analysis". A large body of literature is accumulating on the methods for meta-analysis that can be referred to for guidelines on how meta-analyses of observational epidemiologic studies should be performed.[17] Meta-analysis should be considered a systematic method to aid in evaluating scientific evidence in an area. It allows quantitative evaluation of explanations for differences in results. Indeed, it should be viewed primarily as a comparative (analytic) method, rather than simply as a means for providing an overall quantitative summary or "synthesis" of the study results.

The principle also refers to the preference of obtaining individual study data and combining these for a "pooled" analysis. One should recognize, however, that pooling raw data from all published and unpublished studies is often impossible, and when possible is a time-consuming and labor-intensive process. The principle also refers to "exposure-disease" relations in order to encompass a wider range of health risk scenarios. The principle lists the main components that should be included in a meta-analysis. A protocol for the meta-analysis should include several specific steps. These steps are:

1) *clearly identifying the study variables (i.e., disease outcome, exposure, confounders, effect modifiers, intermediate factors on the causal pathway);*

2) *identifying all studies performed;*

---

[17] See Greenland S., Chapter 32 - "Meta-analysis", In: Rothman KJ, Greenland S., *Modern Epidemiology* (Second Ed. Philadelphia: Lippincott-Raven Publishers, 1998).

3) identifying and extracting the relevant information from those studies;

4) quantifying the effects using the appropriate analytic models;

5) assessing heterogeneity of results across studies and possible sources of heterogeneity, including biases;

6) presenting the individual study results and, where appropriate, summary estimates across studies using descriptive, graphical and statistical analyses;

7) performing sensitivity and influence analyses to evaluate the assumptions made in the meta-analysis; and

8) interpreting the results with consideration of the limitations of the individual studies and the meta-analysis itself.

Two main objectives of meta-analysis are specifically mentioned in this principle. First is the analytic objective, which is the thorough assessment of differences across studies; this objective is essential unless there are absolutely no conflicts among the study results. Second is the synthetic objective, in which a summary estimate of association is constructed; this objective is justifiable only if there is little conflict among the study results used to create the summary*.**

*III.A. RECOMMENDED ADDITIONAL PRINCIPLES
FOR THE USE OF EPIDEMIOLOGIC DATA
IN HEALTH HAZARD IDENTIFICATION
AND RISK ASSESSMENT*

1. Data from well-conducted epidemiologic studies should be given more weight than data from animal or in vitro experimentation. Animal and in vitro studies, however, may provide clues regarding the mechanism underlying associations and aid in understanding causality. Epidemiological studies without biological plausibility should be considered less definitive.

2. Epidemiologic data and study reports, and their impact on the overall risk assessment, should be evaluated by epidemiologists. Scientists from other relevant disciplines should be consulted regarding consistency between the epidemiologic data and non-epidemiologic data.

3. The potential for study results being the result of chance, bias, or confounding should be carefully considered and explored in the assessment.

4. Reasons for all significant inconsistent or implausible results in study findings should be explored. The Bradford Hill factors (or the Surgeon General's criteria) may provide a useful framework for assessing whether there is sufficient evidence of a causal relationship.[20] A well-conducted statistical meta-analysis (consistent with revised Principle B-6), or a careful qualitative meta-analysis, can provide a useful means for exploring the degree of inconsistency or consistency in the available studies in connection with consideration of Hill's factors.[21]

---

[20] Hill's factors and the Surgeon General's criteria are explained and discussed in Appendix B of the London report.

[21] Several of the epidemiologists noted their wariness towards use of Hill's factors due to their many limitations (see, *e.g.,* the discussion at pp. 24-28 in the chapter on "Causation and Causal Inference" by K.J. Rothman and S. Greenland in MODERN EPIDEMIOLOGY (2d ed. 1998, Lippincott-Raven)), and the inappropriately rigid manner in which those factors or criteria had been applied by some; and they

5.   Statistical methods for adjusting study findings should not be employed unless they have been sufficiently validated and are consistent with available relevant biological and toxicological information on mechanisms.

6   All relevant data and study findings should be considered. If there are significant gaps in the reported data or findings, they should be addressed by first consulting with the investigators, and then by reflecting uncertainties in the assessment as necessary.

7.   The suitability and sufficiency of the epidemiologic data for evaluating the specific exposure circumstances of concern to risk managers and the public should be discussed and reflected in the overall assessment.

8.   In a quantitative risk assessment, any significant uncertainties in the hazard, exposure, or dose-response assessments should be reflected in the overall results. The overall results should include information on dose-response, consistency or inconsistency between human and animal data, and the ability to estimate a likely threshold for adverse effects in humans.

---

indicated their preference for use of the more detailed analysis outlined in their revised version of the London principles. See also the limited discussion of some of these issues above in the Discussion Summary above at pages 32-33.

9. The potential importance of the public health issues under consideration should not be a rationale for giving more weight to epidemiologic findings that are otherwise less certain.[22]

10. Hazard identification and risk assessment results should be communicated in a manner that addresses the likely concerns of risk managers and exposed populations.

III.B. QUESTIONS FOR SUPERVISING RISK ASSESSORS AND RISK MANAGEMENT OFFICIALS TO ASK THE EPIDEMIOLOGISTS AND OTHER RISK ASSESSMENT TEAM MEMBERS

Preamble

1.1. The following questions assume that an interdisciplinary team of reviewers, including a number of expert epidemiologists, has examined the relevant epidemiologic evidence (and that scientists from other disciplines have reviewed any relevant animal or other data) and prepared at least a preliminary risk assessment, including a risk characterization, the latter of which has been read by the person(s) asking these questions.

1.2. If there are specific legal criteria that are applicable to a particular risk assessment issue, the questions below should be tailored to those criteria.

1.3. These questions are not necessarily the final step in the risk assessment process, since it is assumed that the questions may

---

[22] This factor can be considered appropriately during the risk management phase of a regulatory decisionmaking process..

expose areas that require additional work or clarifications, resulting in revisions and another similar review.

1.4. It is assumed that a hazard identification or risk assessment document is being prepared in order to inform the general public and/or to determine whether government regulations are needed, and what regulations, if any, should be promulgated.

1.5 Unlike the other two sets of questions (the first for evaluating individual studies, and the second for evaluating a body of studies), these questions are not all worded so that a Yes answer is preferable. This is intentional.

Questions/Checklist Regarding the
Weight of the Human Evidence

2.1. Will the assessment team be regarded as having a high degree of relevant expertise and absence of bias?

2.2. Were any policy or value preferences put aside, and will the assessment be viewed as scientifically objective?

2.3. Are there sufficient data of sufficient quality to draw firm conclusions, or are there significant uncertainties?

2.4. Were there any significant disagreements among the team within their areas of expertise?

2.5. Are there any aspects of the analysis that would not be regarded as meeting accepted scientific norms? Are you confident that all significant aspects of the assessment would be accepted by an expert external peer review panel?

2.6. Is there sufficient epidemiologic evidence presented in the assessment to show convincingly a causal relationship between

56

exposure to the agent and an adverse health effect, or is the evidence of lesser weight?

2.7. *What additional work, if any, would need to be done to address and resolve any significant uncertainties? Is such work feasible? How much time and money would be required? Are there studies under way that could resolve significant uncertainties?   How long will it be before their findings are available?*

2.8. *More specifically, with regard to assessing the weight of the evidence --*

a. *Does the assessment take into consideration all studies and data that should be considered?*

b. *Were there studies or data that were not taken into consideration, or which should be given limited weight or relevancy, due to uncertainties concerning their possible biases or other limitations?*

c. *If any arguably significant studies or data were not taken into consideration, what was the rationale? Were they seriously flawed in some way? Were there problems in determining their relevancy to the risk assessment situation under consideration?*

d. *Were some studies given more weight than others; and if so, is the rationale clearly expressed and scientifically valid?*

e. *Are there uncertainties in the data or study reports that might be explained by obtaining additional information from the investigators, and has an attempt been made to do so?*

f.   *Has the nature of the exposure been properly and consistently characterized in the assessment, or does the assessment need to differentiate in some manner among different exposure circumstances?*

g.   *Are the study findings within the overall body of evidence consistent?*

h.   *If the study findings are not consistent, can the inconsistencies be explained?*

i.   *Have plausible alternative explanations for associations been explored?*

j.   *Have any statistical adjustments made for potential confounding or bias been validated, and have they taken into account all of the likely potential sources of confounding and bias?*

k.   *Are the study findings relied upon for the conclusions in the assessment sufficiently strong to rule out effectively the possibility that an apparent association is due to chance, bias or confounding that cannot be sufficiently identified and adjusted for?*

l.   *Have any temporality issues been addressed? (i.e., Is it clear that any adverse health effects developed a sufficient length of time after exposure, rather than having existed or been developing prior to exposure?)*

m.   *Are there biological or other data that are not coherent with the epidemiologic findings, and have these issues been addressed?*

n.   *Are there biological data that argue strongly for or against the plausibility of a causal relationship?*

o.   *Should we expect the studies to show a dose-response relationship, and do they?*

p.   *Are there sufficient exposure data of sufficient accuracy to support a reasonable dose-response assessment?*

q.   *Are the findings from the epidemiologic studies consistent with relevant animal data and any quantitative assessment method(s) employed in the assessment, particularly in the area of dose-response? Are there reasons for regarding animal data as not significant or of limited value for supplementing the human evidence?*

r.   *If there are not sufficient data for establishing a dose-response relationship, are there sufficient data to determine a no-observable-adverse-effect level (NOAEL)?*

s.   *Have any significant uncertainties (including in the hazard assessment) been reflected in the quantitative estimate(s) of risk (if any)?*

2.9.   *Is it clear from the risk assessment document how the data were evaluated, particularly with regard to any issues likely to be considered significant or controversial? Does the document describe the analytic methods sufficiently to allow other scientists to replicate and check the analysis?*

2.10.   *If you were given responsibility for challenging the risk assessment, what particular points would you raise? Have those points been clearly and satisfactorily addressed in the assessment document?*

2.11. Are there certain caveats, qualifications, or explanations that should be added to the findings to ensure that regulatory officials and the public understand their relevance? For example, do the findings appear to apply only to certain subpopulations or under certain conditions?

2.12. Upon further consideration, are there clarifications or other revisions that should be made to the risk characterization?

2.13. What questions that have not been asked should be asked?

2.14. In the end, what level of confidence do you have in your overall conclusions and the supporting explanations in the assessment?

PUBLICATIONS USED IN THE REVIEW

Animal and Human Biological Studies

Russo J, Russo IH, "Susceptibility of the mammary gland to carcinogenesis. II. Pregnancy interruption as a risk factor in tumor incidence", *Am. J. Pathol.* 100:497-512 (1980).

Russo J, Rivera R, Russo IH, "Influence of age and parity on the development of the human breast", *Breast Canc. Res. Treat.* 23:211-18 (1992).

Salazar H, Tobon H, "Morphologic changes of the mammary gland during development, pregnancy and lactation", In: Josimovihc J. ed., *Lactogenic Hormones, Fetal Nutrition and Lactation*, 221-77 (J Wiley & Sons, NY, 1974).

Whittaker PG, Stewart MO, Taylor A, Lind T, "Some endocrinological events associated with early pregnancy failure", *Br. J. Ostet. Gynaecol.* 96:1207-14 (1989).

Witt BR, Wolf GC, Wainwright CJ, et al. "Relaxin, CA-125, progesterone, estradiol, Schwangerschaft protein, and human chorionic gonadotropin as predictors of outcome in threatened and nonthreatened pregnancies", *Fertil. Steril.* 53:1029-36 (1990).

Yen SSC, "Endocrinology of pregnancy", In: Creasy RK, Resnik R, eds., *Maternal-Fetal Medicine: Principles and Practice* (WB Saunders, Philadelphia, 1989).

Assigned Epidemiologic Study Reports

Adami H-O, Bergstrom R, Lun E, Meink O, "Absence of association between reproductive variables and the risk of breast cancer in young women in Sweden and Norway", *Br. J. Cancer* 62:122-26 (1990).

Daling JR, Malone KE, Voigt LF, White E, Weiss NS, "Risk of breast cancer among young women: relationship to induced abortion", *J. Nat. Cancer Inst*. 86:1584-92 (1994).

Daling JR, Brinton LA, Voigt LF, Weiss NS, Coates RJ, Malone KE, Schoenberg JB, Gammon M, "Risk of breast cancer among white women following induced abortion", *Am. J. Epidemiol*. 144(4):373-80 (1996).

Howe HL, Senie RT, Bzduch H, Herzfeld P, "Early abortion and breast cancer risk among women under age 40", *Int. J. Epidemiol*. 18:300-04 (1989).

La Vecchia C, Negri E, Franceschi S, Parazzini F, "Long-term impact of reproductive factors on cancer risk", *Int. J. Cancer* 53:215-19 (1993).

Laing AE, Demenais FM, Williams R, Kissling G, Chen VW, Bonney GE, "Breast cancer risk factors in African-American women: The Howard University Tumor Registry experience", *J. Natl. Med. Assoc*. 85:931-99 (1993).

Lipworth L, Katsouyanni K, Ekbom A, Michels KB, Tricholpoulos D, "Abortion and the risk of breast cancer: a case-control study in Greece", *Int. J. Cancer* 61:181-84 (1995).

Melbye M, Wohlfahrt J, Olsen JH, Frisch M, Westergaard T, Helweg-Larsen K, Andersen PK, "Induced abortion and the risk of breast cancer", *New Eng. J. Med*. 336(2):81-85 (1997).

Nishiyama F, "The epidemiology of breast cancer in Tokushima prefecture", *Shikoku Ichi* (*Shikoku Medical Journal*) 38:333-43 (1982) (translation from Japanese).

Epidemiologic Study Reports with Emphasis on Recall/Response Bias (first two assigned)

Lindefors-Harris B-M, Eklund G, Adami H-O, Meirik O, "Response bias in a case control study: Analysis utilizing comparative data concerning legal abortions from two independent Swedish studies", *Am J. Epidemiol.* 134:1003-08 (1991).

Rookus MA, van Leeuwen FE, "Induced abortion and risk for breast cancer: Reporting (recall) bias in a Dutch case-control study", *J. Nat. Canc. Inst.* 88(23):1759-64 (1996).

Jones EF, Forrest JD, "Underreporting of abortion in surveys of U.S. women: 1976 to 1988", *Demogr.* 29:113-26 (1992).

Assigned Statistical-Qualitative Meta-Analysis

Brind J, Chinchilli VM, Severs WB, Summy-Long J, "Induced abortion as an independent risk factor for breast cancer: A comprehensive review and meta-analysis", *J. Epidemiol. & Comm. Hlth* 50(5):481-96 (1996).

(see also the Michels & Willett qualitative review, below)

Unassigned Epidemiologic Study Reports

Andrieu N, Clavel F, Gairard B, et al., "Familial risk of breast cancer and abortion", *Cancer Detect. Prev.* 18:51-55 (1994).

Bu L, Voigt L, Yu Z, Malone K, Daling J, "Risk of breast cancer associated with induced abortion in a population at low risk of breast cancer", *SER Abstracts* S85 (1995).

Brinton LA, Daling JR, Liff JM, et al., "Oral contraceptives and breast cancer risk among younger women", *J. Natl. Cancer Inst*. 87:827-35 (1995).

Brinton LA, Hoover R, Fraumeni JF Jr., "Reproductive factors in the aetiology of breast cancer", *Br. J. Cancer* 47:757-62 (1983).

Burany B, "Gestational characteristics in women with breast cancer", *Jugosl. Ginekol*. Opstet. 19:237-47 (1979) (in Serbo-Croatian).

Dvoyrin VV, Medvedev AB, "Role of women's reproductive status in the development of breast cancer", In: "Methods and progress in breast cancer epidemiology research", Oncology Science Center of the USSR Academy of Sciences" 53-63 (Tallin, 1978) (translation from Russian).

Hirohata T, Shigematsu T, Nomura AMY, Nomura Y, Horie A, Horohata I, "Occurrence of breast cancer in relation to diet and reproductive history: A case-control study in Fukuoka, Japan", *Nat. Cancer Inst. Monogr*. No. 69, 187-90 (1985).

Laing AE, Bonney GE, Admas-Campbell L, et al., "Reproductive and lifestyle factors for breast cancer in African-American women", *Genet. Epdemiol*. 11:A300 (1994) (abstract -- apparently a follow-up to the Laing et al. 1993 study report listed above under assigned study reports).

Le M-G, Bachelot A, Doyon F, Dramar A, Hill C, "Oral contraceptive use and breast or cervical cancer: Preliminary results of a French case-control study", In: Wolff J-P, Scott JS, eds., *Hormones and Sexual Factors in Human Cancer Aetiology* (Elsevier, Amersterdam, 1984).

Moseson M, Koenig KL, Shore RE, Pasternack BS, "The influence of medical conditions associated with hormones on the risk of breast cancer", *Int. J. Epidemiol*. 22:1000-09 (1993).

Newcomb PA, Storer BE, Longnecker MP, Mittendorf R, Greenberg ER, Willett WC, "Pregnancy termination in relation to risk of breast cancer", *JAMA* 275:283-87 (1996).

Rosenberg L, Palmer JR, Kaufman DW, Strom BL, Schottenfeld D, Shapiro S, "Breast cancer in relation to the occurrence and time of induced and spontaneous abortion", *Am J. Epidemiol*. 127:981-89 (1988).

Segi M, Kukushima I, Jujisaku S, et al., "An epidemiological study on cancer in Japan", *GANN* (The Japanese Journal of Cancer Research) 48(Supp):1-63 (1957) (selected pages only).

Watanabe H, Hirayama T, "Epidemiology and clinical aspects of breast cancer", *Nippon Rinsho* 26:1843-49 (1968) (in Japanese, with handwritten translation of some pertinent portions supplied by J. Brind).

Unassigned Review Articles

Michels KB, Willett WC, "Does induced or spontaneous abortion affect the risk of breast cancer?", *Epidemiol*. 7(5):521-28 (1996).

Remennick LI, "Induced abortion as cancer risk factor: A review of epidemiological evidence", *J. Epidemiol. Comm. Hlth* 44:259-64 (1990).

Rosenberg L, "Induced abortion and breast cancer: More scientific data are needed", *J. Nat. Cancer Inst*. 86:15-16 (1994).

Wingo PA, Newsome K, Marks JS, Calle EE, Parker SL, "The risk of breast cancer following spontaneous or induced abortion", *Cancer Causes and Control* 8:93-108 (1997).