Preliminary Technical Review of the Hanford Thyroid Disease Study Draft Final Report

Fred Hutchinson Cancer Research Center

CDC
CENTERS FOR DISEASE CONTROL AND PREVENTION

Preliminary Technical Review of the Hanford Thyroid Disease Study Draft Final Report

Table of Contents

A. Importance of Peer Review in the Scientific Process

The goal of the Hanford Thyroid Disease Study is to identify any possible link between past exposures to I-131 from the Hanford facility and the risk of thyroid disease. Because of both the public health implications and the complex scientific issues associated with this study, it is critically important that each step in the assessment be openly and objectively evaluated not only by the affected public, but also by the scientific community. This type of technical evaluation, often called peer review, is a critical component in the scientific process. Peer review ensures that the scientists conducting the study apply methods that are appropriate to the problem being investigated, it promotes the free exchange and discussion of ideas, and often leads to improvements in both the analysis and the presentation of the study data. In addition, review of the study methods and results by knowledgeable, yet objective, experts can increase the credibility of the work within both the scientific and public communities. Usually, to ensure objectivity, the peer review process is confidential with unknown reviewers evaluating the work of unidentified investigators. However, since its inception, the Hanford Thyroid Disease Study has been conducted with complete openness to public scrutiny. As a result, the Centers for Disease Control and Prevention (CDC) is dedicated to continuing this openness throughout the review process for the draft Hanford Thyroid Disease Study Report.

B. Preliminary Technical Review of the Hanford Thyroid Disease Study Draft Final Report

The CDC received the Hanford Thyroid Disease Study Draft Final Report from the Fred Hutchinson Cancer Research Center (FHCRC) on September 30, 1998. Shortly after that, copies were also sent to a group of scientists who were asked to provide their opinions on the methods, analysis and interpretation of results summarized in the draft report. The CDC, together with the study's principal investigators from the Fred Hutchinson Cancer Research Center, selected the scientists to be used as reviewers based on their recognized expertise in areas such as diseases of the thyroid, radiation epidemiology, and statistics. In addition, a group of scientists from within CDC was also asked to review the draft report. The reviewers were asked to study the draft report and share their opinions with the study team, the CDC and each other at a meeting held in Atlanta on November 13, 1998. In addition, reviewers were asked to provide CDC with written summaries of their evaluation of the methods, analysis and presentation of results described in the draft report.

On page 4 of this document you will find will find a list of the scientists asked to provide technical reviews of the draft report. Verbatim copies of the written comments provided to CDC by the reviewers are provided on the following pages along with the minutes of the November 13, 1998 meeting of the reviewers. Two important ideas should be kept in mind as you examine these reviews. First, the comments submitted by this group of technical reviewers are only the first of many CDC expects to receive concerning the draft report, both from affected citizens and the scientific community. For example, a committee of The National Academy of Sciences, comprised of experts in a variety of scientific disciplines, will review the draft report in early

February 1999. This committee will spend about 2 to 3 weeks producing a written evaluation of the Hanford Thyroid Disease Study, which will be available to the public when completed.

Other persons who would like to comment on the Hanford Thyroid Disease Study Draft Final Report are encouraged to do so by writing to

Mr. Michael Donnelly
Radiation Studies Branch
MS F-35
Centers for Disease Control and Prevention
4770 Buford Highway NE
Atlanta, GA 30341.

In order to begin work on the final Hanford Thyroid Disease Study report, we request that comments by sent by July 1, 1999. All comments will be given consideration in preparation of the final report.

A second important point to remember when examining these review comments is that no changes have yet been made to the draft report based on the comments of the technical reviewers listed in this document. The draft report released today is the same as the report received by the CDC on September 30, 1998. The goal in providing copies of the initial reviewers' comments on the report is to illustrate the public and scientific review process CDC expects the draft report to undergo and to continue the emphasis on complete openness in both conducting and reporting the results of the Hanford Thyroid Disease Study.

C. Participants in the Preliminary Technical Review of the Hanford Thyroid Disease Draft Final Report

Reviewers from Outside the Centers for Disease Control and Prevention

David V. Becker, M.D.
New York Presbyterian Hospital-
Cornell University Medical Center
525 East 68th Street
Room Starr 221
New York, NY 10021

Maureen Hatch, Ph.D.
Department of Community Medicine
Mount Sinai School of Medicine
1 Gustave L. Levy Place
Box 1043
New York, NY 10029

Laurence Needleman, M.D.
Thomas Jefferson University
Division of Ultrasound
Main Building--Seventh floor
132 South Tenth street
Philadelphia, PA 19107-5244

Arthur Schneider, M.D., Ph.D.
University of Illinois
835 S. Wolcott
Room E-625
Chicago, IL 60612

Centers for Disease Control and Prevention Reviewers

Owen Devine, Ph.D.
Christie Eheman, Ph.D., M.S.H.P.
Dana Flanders, M.D., Sc.D.
Paul Garbe, D.V.M., M.P.H.
Charles Miller, Ph.D.
David Olson, Ph.D.
Judith Qualters, Ph.D.

D. Comments of Technical Reviewers on the Hanford Thyroid Disease Study Draft Final Report

The following are verbatim copies of comments on the Hanford Thyroid Disease Study Draft Final Report submitted by the technical reviewers.

Comments of David V. Becker, M.D.

Dr. Becker is Professor of Radiology and Medicine at the Cornell University Medical Center. He has been the co-investigator in a number of epidemiologic studies of thyroid disease and has collaborated with the National Cancer Institute on studies of thyroid disease resulting from radiation exposure related to the Chernobyl disaster. In 1992, Dr. Becker participated in the CDC sponsored workshop on thyroid ultrasound that resulted in ultrasound examinations being included in the Hanford Thyroid Disease Study.

DRAFT

REPORT FOR CENTERS OF DISEASE CONTROL

David V. Becker, M.D.
New York Presbyterian Hospital-Cornell Medical Center
December 10, 1998

The Hanford Thyroid Disease Study has the well defined objective of determining whether there was any increase in thyroid disease in the exposed population that could be attributed to I-131 released into the atmosphere from the Hanford Nuclear Site between 1944 and 1957. Over 5,000 individuals were included in the study cohort and of these, 84% were located, a remarkably successful follow-up effort 40 years after the original exposure. It was planned to correlate any increase in thyroid disease detected to the estimated radiation dose in order to describe a dose-response relationship. Other objectives included a determination of whether hyperparathyroidism had increased in those exposed.

It would appear from the draft report that the study has been carried out in a careful and effective way following a well designed experimental protocol. The defined population to be studied appears appropriate and a major effort at many different levels was rewarded with remarkably successful location and follow-up. The study cohort was identified on the basis of proximity to the Hanford site in what appears to be a thoughtful and well carried out dose reconstruction (dosimetry) effort which developed what seems a credible approximation of individual thyroid radiation dose. A computerized interview technique was used to assist in obtaining an estimate of the dietary intake and residence data needed for individual dose reconstruction. A special effort was made to approach the selected study participants in an open and sympathetic way and the investigator's efforts were rewarded with 84% of the population agreeing to participate. Since

accurate dose estimation is critical to the study, thyroid radiation dose was estimated not only by primary algorithms but also multiple variations, an effort which if not exhaustive at least provided assurance of maximum return in obtaining valid dose estimates.

For most categories of thyroid disease, the data produced showed no evidence of a relationship between estimated thyroid radiation dose from Hanford fallout and the cumulative incidence of specific thyroid abnormalities. These important results are credible and convincing.

I would like to comment on a few issues that might profit by some further discussion in the report. It is surprising that the presumably normal study population had such a large number of prior diagnostic X-ray exposures. That 36% of the population had prior upper GI series and 34% had X-rays of the head, to say nothing of the 24% of the population that had a CT scan of the upper body raises a question as to whether the selected study population is truly representative. The availability of an appropriate control population for comparison would add another layer of reassurance. An informal and anecdotal query of a limited number of diagnostic radiologist colleagues suggests that these numbers are beyond what might be expected in ordinary practice with a “normal” population.

It is of interest and a little disturbing that mortality in the cohort was 20% higher than expected. The primary contributor to this number was apparently congenital abnormalities and perinatal problems, which amounted to an excess of over 2 times that, expected. This could hardly be a radiation effect since many of these births occurred before Hanford went into operation but further discussion would be helpful.

The population also had a surprising amount of thyroid disease although its prevalence was not dose related. The overall incidence of almost 19% autoimmune thyroiditis with this number reaching 24% for women in the study is more than might be expected from results of normal population studies. The numbers for hypothyroidism (17% of the total population, 27.5% of the women) is also higher than the one might expect from other epidemiologic studies of presumably normal populations. However, the ultrasound studies showed that 46.5% of the population had one or more ultrasound detected abnormalities, a number that correlates very closely with the pathologic findings of Mortensen (1955) where an autopsy study showed a similar number of gross and microscopic abnormalities in a non-exposed midwestern population. These findings raise questions that should be answered with regard to the population studied although it is hard to see how significant bias could be introduced considering the way in which the population was selected.

The introduction of ultrasound into the study, I believe, greatly solidified its results. Nonpalpable ultrasound detected abnormalities seems a reasonable category and the use of ultrasound in combination but separate from palpation in demonstrating discrete masses and defining nodules is important for comparison to a large body of epidemiologic data based upon palpable findings. This study data should be extremely useful to evaluate the findings on ultrasound which to date have not really been studied in a major epidemiologic study with both modalities. At the same time, it would be particularly useful if repeat ultrasound studies could be carried out on this same

population to assess the life history and development of the non-palpable ultrasound defined thyroid abnormalities. Although clinical practice findings suggests that many of these may appear and disappear in the course of the patient's life, any controlled follow-up information that could be developed would be of major clinical importance.

A methodological quality assurance issue that should be discussed is the importance of review of FNA cytologic diagnoses- which can be done retrospectively. There is also some uncertainty about the practice of assigning a diagnosis to FNA samples with insufficient cells for examination.

In summary, I find the study well conceived and carefully carried out. The findings are credible and, in general, within the range of what might be found in other population studies. Because the absence of any correlation with radiation dose is so important, perhaps further review of the dosimetry estimations could be carried out. It would be particularly helpful in evaluating the results if a comparison reference population could be developed for assistance in analysis of some of the data, as noted above. However, on the whole, the study is a coherent, well developed, and effective epidemiologic study of a very difficult situation.

Comments of Maureen Hatch, Ph.D.

Dr. Hatch is Professor of Community Medicine at the Mount Sinai School of Medicine. She has participated in a number of investigations of the health effects resulting from the Three Mile Island accident in 1979. Dr. Hatch has been a member of the Hanford Thyroid Morbidity Advisory Committee since 1995.

DRAFT

TO: Paul Garbe
FROM: Maureen Hatch
RE: Hanford Thyroid Disease Study Draft Report
DATE: December 10, 1998

As agreed at the meeting, my comments will focus on the mortality analysis.

1. The SMRs were calculated using Washington State as the standard. Are there any regional data available that would provide a better comparison or are the statewide data right?

2. Two findings of the mortality analysis are especially notable; the excess in perinatal mortality and in fatal congenital anomalies, and the particularly high mortality among those born in Franklin County.

   -To what extent do these overlap: ie, is the excess in Franklin County due to perinatal mortality? This should be addressed.

3. Re: the birth year category analysis--as I said at the meeting, year of birth as an indicator of exposure can be somewhat misleading when one is dealing with congenital anomalies or pregnancy complications. Exposure at conception or in early pregnancy may be relevant, and this might be in the previous calendar year.

   -If you have the actual birth dates, you could estimate year of conception and examine the data that way. This would be a more appropriate analysis.

4. As agreed at the meeting, the particular perinatal conditions and congenital defects should be described. This could prove informative, especially if coupled with data on place of birth and year of conception.

5. Again, as per the meeting, these results should be discussed in the context of other relevant studies such as Sever et al. (Am J Epidemiol 1988; 127). Even though HTDS wasn't designed to explore mortality findings in detail, they deserve more consideration than is currently given on p.8 of Section IX; for example, one could suggest appropriate follow-up studies.

6. In terms of the excess in cardiovascular mortality, are there any obvious differences in risk factor prevalence that would explain it?

_____

Overall, the investigators have done an excellent job. Addressing the major comments from the November meeting would make the Report truly outstanding. The major points, in my view, are the following:

- Undertake a review of FNA cytology even if it means a few months' delay. The fact that several reviewers in addition to Dr. Schneider feel this is important speaks volumes.

- Provide a brief discussion of statistical issues like 'significance' and multiple comparisons.

- Include more user-friendly data approaches in addition to dose-response.

- Pay more attention to trends in the data that may give rise to public concern.

Comments of Charles Miller, Ph.D.

Dr. Miller is Chief of the Dosimetry Section of the Centers for Disease Control's Radiation Studies Branch. He is an expert in the area of mathematical modeling of how contaminants, such as radionuclides, move through the environment. In addition, Dr. Miller is a member of the National Council on Radiation Protection.

To:	Paul Garbe, D.V.M.
From:	Charles W. Miller, Ph.D.
Date:	November 30, 1998
Subject:	Review of the Hanford Thyroid Disease Study Draft Final Report

Per your request, I have reviewed the report listed above. In accordance with my knowledge and expertise, I have focused on the radiation dosimetry portions of this report in my review. In general, I find that the dosimetry considerations are very reasonable and appropriate for this study. The authors are to be commended for including an analysis of alternate methods for classifying dose in addition to the primary methodology. Examination of multiple dose classification procedures greatly strengthens the argument that a significant positive dose response is not supported by the data.

A few specific comments, all taken from my review of Section V, are presented below for the authors' consideration.

P. 87 The study uses the information from the study subject about water source to determine whether or not the pasture used by the appropriate milk cows was irrigated or not. What is the basis for using this indirect information in this manner? What would have been the impact on dose of misclassifying the cow•s source of pasture?

P. 87 The authors state that in April 1996 they brought to the attention of the HEDR Task Completion Working Group and others the issue of inconsistencies in dairies between the HEDR data and the study subjects. Has there been a resolution of this issue at this time? How did HTDS handle this issue in the dose calculations? What impact does this issue have on the dose estimates?

P. 88 The authors state that •... it was impractical to allow a participant's reference dose category to change over time.• What was the impact of this computer code limitation on the calculated doses, and thus the dose response analysis?

P. 90 The authors indicated in Section E.3.c that doses from the Nevada Test Site were calculated for the study subjects. There will be some reviewers who will argue that these doses should have been added to the Hanford doses and incorporated into the final dose response analysis. This issue is addressed in Section VII.D, but it needs to discussed here, too, or the later section where it is discussed referenced here.

P. 116 In this section it is stated that •...four participants were determined to be non-evaluable.• On p. 118 the number of non-evaluables is given as six. I don•t know which number is correct, but I believe that the same number should be given in both of these places.

P.118 What is the impact on the dose estimates of using •fuzzy date codes• for residences, and how did you evaluate this impact?

You requested my review comments by the end of November, and hence I am providing these comments to you. However, it should be noted that one significant section is still missing from the report. A discussion of the impact on the dose response analysis of the uncertainty in the dose estimate is a necessity for this report to be complete. I have seen a brief verbal presentation on this subject, and I understand that written material is currently being prepared. I do not consider my review to be complete, however, until I have had an opportunity to examine this missing material.

I hope the comments above are useful to both you and the authors. Please contact me if you have any questions concerning them.

Comments of Laurence Needleman, M.D.

Dr. Needleman is Associate Professor of Radiology, Thomas Jefferson University. He is an expert in the use of ultrasound and fine needle aspiration in the diagnosis of thyroid diseases. Dr. Needleman participated in the 1992 CDC workshop that led to inclusion of ultrasound examinations in the Hanford Thyroid Disease Study design.

My comments will largely confine themselves to my two areas of expertise: ultrasound and thyroid fine needle aspiration.

The report makes clear the purposes served by the introduction of thyroid ultrasound (G.3.b). One of these is that ultrasound may show non palpable abnormalities and another is that it provides an objective record of the characteristics of an "abnormalitiy." Size of a lesion is an important factor but this is not stressed in the report.

FNA was performed on those with a discrete, palpable nodule or dominant nodule in a multinodular gland. Starting in 1994 FNA was expanded to those patients with nonpalpable ultrasound nodules greater than 1.5 cm. The choice of FNAing lesions greater than 1.5 cm. is a strength of the study but it is not stressed in the report. This size is small enough to indicate the investigators were aggressively pursuing thyroid disease. Choosing a smaller threshold is unrealistic given the large number of tiny lesions that are frequently found in ultrasound scans of patients in this age range, and their uncertain significance.

There is no sense of how many nodules greater than 1.5 cm. were biopsied or unbiopsied (i.e. ultrasound detected, nonpalpable nodules that remained unbiopsied). The report does not lay out the results of the FNA data as clearly as it might. The total number of lesions greater than 1.5 could be determined and it can be shown how many underwent needle aspiration and what the results were. The followup of those unbiopsied (e.g. results of ultrasound follow-up program, post-clinic medical records review) could be presented and the number of unevaluated lesions stated.

These data can also be presented starting with the nodules that were biopsied and determine how they were discovered (palpation, ultrasound) and show the pathological results and final assessment.

The FNA results are somewhat unclear in the draft report. For instance, on page 8 of the executive summary it states that 259 participants had FNA. 62 (24%) were recommended to have further biopsy but in the

post-clinic medical records review it states medical records documenting further diagnostic studies were requested for 35 participants (page 9 of executive summary). What happened to the other 27 participants (62 less 35)? Are these the patients who have a nodule which is suspicious for malignancy or neoplasm (Background, page15)? But this number appears to be 16 (the difference between line 2 on Tables VIII-18 and VIII-22 and -23.

There is discussion in the field procedures about quality control for the radiologists and the sonographers (Field procedures page 107 to 109). The radiologists had a high level of concordance and this is clearly stated in the report. The sonographers had somewhat greater differences including discrepancies in number of nodules and presence or absence of nodules greater than 5mm. These discrepancies are not addressed and should be commented on briefly. In my opinion, agreement for the presence or absence of a greater than 5mm nodule is acceptable (approximately 10% of cases). The nodule size selected was 5 mm. which is quite strict and is of a size where a nodule and background heterogeneity of the gland texture may overlap. Had the sonographers quality been judged on a larger nodule size, on the order of 10 mm., I suspect the discrepancies would be far less. This may be important in the interpretation of dose response for UDAs. The dose response was determined for the presence or absence of UDAs. The QC results support this.

It may be desirable to do another dose response analysis using the number (or size and number), and not just the presence of UDAs. This may not be justified for all lesions greater than 5 mm. since there appears to be larger intersonographer differences determining the number of these lesions. A reanalysis of the sonographers may show better concordance for larger lesions. A dose response for number of lesions might be possible for larger lesions, e.g. greater than 10 mm. but not for greater than 5 mm. lesions.

This is potentially an important question given the results when certain geostrata were removed from the analysis a p value less than .05 was found (Results, page 86).

It may be that the UDAs do not reflect thyroid disease but may show a dose response, perhaps in the way some people chronically tanned may have different skin characteristics than those who stay out of the sun. The presence of wrinkles or other stigmata may not be a disease but the experienced dermatologist may be able to determine if the patient had a lot or little sun exposure. By analogy more (or a larger volume of) nodules may be more likely to be present with those exposed to higher

doses without indicating thyroid disease.

The study is rightfully uncertain what to recommend to patients with UDAs. Nonetheless they state in the Recommendation for Health Care Providers (Appendix of the draft final report) that the number and size of nonpalpable abnormalities "likely influences the probability that thyroid disease is present and should be considered in the management of that patient."

The study should consider if a dose response analysis of the number and size of UDAs is feasible and should be performed (taking into account my caveat regarding sonographer concordance).

There is little or no quality control on the pathology results. This is a potential source of criticism. Rereading of some of the cases may be in order.

The FNA technique or pathologist may be called into question because the report does not discuss the FNAs in detail. Sixty two of 259 FNA (24%) were recommended to have further biopsy or surgery. While this at first may seem high, the report does not show that this number may be reasonable. Twelve of the participants had thyroid cancer and 14 follicular adenoma. This leaves 13% of the FNAs needing further evaluation. This still seems high to me but may be lower if there are other cases that could be explained away. The report should go through these cases and show what percent of cases actually needed confirmation due to the uncertainty of the FNA/pathologist. They should compare the nondiagnostic percentage in this study to published thyroid FNA results.

Finally, I have some questions. This is somewhat outside of my expertise and I am asking not as my role as ultrasound expert but as an interested physician asked to comment when the report is not clear or to predict questions which may be raised when the study is presented.

I am not clear that some of the "nodule definitions" used in this study correspond to the "neoplasia" which Kerber et al used. The discussion talks about thyroid neoplasia (Discussion, page 15) and the Utah study, but are the results comparable? Clearly carcinoma was analyzed in the two studies but were benign neoplasms analysed separately in this study? In this study wasn't the dose response determined for benign nodules which included other entities besides benign neoplasms? Aren't the benign nodules in this study therefore different than Kerber[1]s benign neoplasms?

Was a dose response analysis of benign neoplasms done?

Lastly, is it reasonable to perform a dose response analysis on different combinations of neoplasms, that is 1) adenomas and carcinomas together, excluding colloid and non-neoplastic nodules and 2) adenomas and carcinomas and nodules suspicious for malignancy, excluding colloid and non-neoplastic nodules?

Comments of David Olson, Ph.D.

Dr. Olson is Chief of the Biometry Branch in the Centers for Disease Control and Prevention's Division of Environmental Hazards and Health Effects. His area of expertise is in the statistical analysis of epidemiologic data.

Date: November 12, 1998
To: Paul Garbe
Re: Review of Hanford Thyroid Disease Study Draft Final Report

I have reviewed this report and following are my responses to the four specific questions to be considered:

The analysis was carefully, meticulously, and in my opinion, appropriately carried out. However, it was not completely carried out. Page 14 of section VII of the report describes an extended analysis approach intended to account for the effects of uncertainty in the dose estimates. I found no evidence in the presentation of results that this was done.

The words used to describe levels of statistical significance in the analysis are sometimes imprecise or contradictory. It may be inferred from the document that the authors intend to define marginal statistical significance as $0.01 <= p < 0.05$ and statistical significance as $p < 0.01$. This needs to be explicitly stated and whether or not these definitions were made a priori. Added adjectives such as barely, clearly, nearly, and only are superfluous. I list a few examples (some OK and some not) from section VIII below, but all descriptions need to be accurate and consistent.

Pages	p-value	Description
43-44	0.045	nearly achieved marginal levels of statistical significance
43-44	0.057	nearly achieved marginal levels of statistical significance
65	0.0026	statistically significant
69	0.039	marginally significant
69	0.011	only marginally significant
71	0.040	barely marginal statistical significance
74	0.059	not statistically significant
76	0.051	barely marginally significant
80	0.050	not significantly higher

Also, on page 90 of section VIII and on page 13 of section IX, mention is made of conducting a large number of significance tests in the context of secondary and alternative analyses and the need for caution in interpreting a specific p-value of 0.003. This caution applies study-wide, not just to this test alone. Can the few significant results observed be explained by chance alone?

The presentation and discussion of results is complete, except as noted above.

In my opinion, the conclusions made are reasonable, assuming that the extended analysis accounting for the effects of uncertainty in the dose estimates yields similar results. Also, the authors make good arguments that results based on just two-thirds of the people originally identified for the study are representative of all. I did not identify flaws in this reasoning.

David Olson

Comments of Dr. Arthur Schneider, M.D., Ph.D.

Dr. Schneider is Professor of Medicine and Chief, Section of Endocrinology at the University of Illinois, Chicago. He has published many scientific articles on the risk of thyroid disease resulting from exposure of the thyroid gland to x-rays. Dr. Schneider has served on the Hanford Thyroid Morbidity Advisory Committee since the beginning of the Hanford Thyroid Disease Study. In addition, he participated in the 1992 CDC workshop that led to inclusion of ultrasound examinations in the Hanford Thyroid Disease Study design.

DRAFT

COMMENTS ON HANFORD THYROID DISEASE STUDY DRAFT FINAL REPORT

Arthur B. Schneider, M.D., Ph.D.

November 16, 1998

I am writing to summarize my evaluation and comments on the report. I reviewed it prior to attending the meeting on November 13, 1998 in Atlanta. During that meeting I felt that I had the opportunity to highlight what I considered to be the strengths of the report as well as the few areas where I felt that it could be strengthened. In writing these comments, I am assuming that my detailed comments will be reflected in the summary of the meeting and that I will have the opportunity to review and confirm them.

My overall impression is that the report is a comprehensive and fair description of a well designed and carefully carried out research project. I believe that the conclusions are appropriate and strongly supported by the data. I believe that the appropriate analyses have been carried out for all of the primary endpoints and these were carefully defined in the study design. I believe that the analyses were carried out completely, the levels of statistical significance were appropriate, the presentation and discussion are complete, the outcomes that include ultrasound were appropriately dealt with, and, as stated above, the conclusions are supported by the findings.

The following are some comments related to the report. They take into account the importance of the report and the wide audience that will use it.

1. The report should emphasize prominently the difference between the primary analyses, formulated as part of the study design, and secondary analyses, conducted after the data have been obtained. In this report, all of the primary analyses demonstrate with high certainty the absence of dose response relationships. There are two types of secondary analyses. The first are analyses carried out to assist in the presentation of the data. The second are analyses looking for other trends or clues to other findings in the data. It

should be made clear that the latter are highly subject to the problem of carrying out multiple analyses and finding, only by chance, that one appears to be significant. In this report this applies to the findings related to serum calcium and small ultrasound detected abnormalities. These may be significant findings or they may be random fluctuations in the data. Further work would be required to determine which is correct. However, since both of these appear to have little clinical implication, it would be hard to justify.

2. Fine needle aspiration cytology. An important strength of the study was that the review of all of the cytology was done by a single cytopathologist, unaware of the radiation status of the subject. The uniformity of interpretation is important. However, now recognizing what the findings are, an additional confirmation is justified. I would consider reviewing all cases of malignancy and a subset of the remaining cases by two or more experienced thyroid cytopathologists. Also, I would ask them to review the cases with hypocellular samples with abundant colloid to determine whether they agree with their classification.

3. Thyroid ultrasound. Thyroid ultrasound was used in a uniform way and entered into the definition of the principal endpoints in a clearly defined way. These definitions combined the physical examination and ultrasound findings and, as appropriate, FNA and pathology results. To confirm the findings of a lack of a dose response relationship, and to investigate further the role of palpation and ultrasound, I suggest carrying out separate analyses for the endpoints of palpable nodules and ultrasound nodules larger than 1.5 cm in average dimension.

4. I found one table in the analysis plan that was not in the final report. This was the number of palpable nodules that were not confirmed by ultrasound. Since the analysis plan was reviewed and agreed to by several groups, it is best to complete all aspects.

5. Presentation of outcomes. I believe that it would be very helpful to have tables that summarize “paths to diagnosis”. I believe that there should be about three of these tables, one for cancer, one for benign, and for all nodules. To explain, by example, using the proposed cancer table, I would have entries such as the following:

   Study diagnosis → Palpable nodule → Ultrasound demonstrated nodule →
   FNA = malignant → Surgical pathology = papillary thyroid cancer.

   Study diagnosis → No palpable nodules → Ultrasound demonstrated nodule →
   FNA = malignant → Surgical pathology = papillary cancer

   Study diagnosis → Palpable nodule → Ultrasound demonstrated nodule →
   FNA = follicular neoplasm →Surgical pathology = follicular carcinoma

   These tables will be helpful to the understanding of how the diagnoses were made. They show how many unusual cases were present. For example, these tables would indicate how often acellular, but colloid rich, fine needle aspirations occurred.

6. I believe that the description of multinodular goiters and multinodular glands and their overlap with the category of thyroid nodules should be made clearer.

7. The laboratory findings should be scanned to be certain that the appropriate units are stated whenever a value is given. Also, while the various TSH methods are described, it would be helpful to describe them by their level of sensitivity.

I hope that you find these comments of assistance. From the onset I have been very impressed by the openness with which this work has been conducted. I realize that in order to maintain this, you are working very hard to prepare this report for open discussion. Please let me know if I can be of any further assistance in your efforts to do this.

E. Minutes of Meeting of Technical Reviewers for the Hanford Thyroid Disease Draft Final Report Held November 13, 1998 at the Centers for Disease Control and Prevention, Atlanta, Georgia

Meeting Agenda

9:45-9:00	Introductions and charge for discussion Drs. Falk, Smith
9:15-10:00	Overview of Draft Final Report Drs. Davis, Kopecky, Hamilton
10:00-11:30	Reviewer discussion with HTDS Investigators Dr. Garbe
11:30-1:00	Working lunch for reviewers
1:00-2:30	Reviewer discussion with HTDS investigators Dr. Garbe
2:30-3:00	Wrap-up Drs. Falk, Garbe, Smith
3:00	Adjourn

Review questions:

Was the analysis carried out appropriately and completely?

Are the levels of statistical significance used in the analysis appropriate?

Are the presentation and the discussion of results complete?

Is the presentation of results on ultrasound appropriate?

Are the conclusions reasonable?

Meeting Minutes

Dr. Garbe convened the meeting at 9:00 am and asked for participants to introduce themselves. He reviewed the agenda for the day and noted the reviewers should offer candid comments on the HTDS draft report. CDC needed both reviewer general comments and suggestions, and specific comments regarding the five questions posed to the review group.

Dr. Davis summarized the history, purpose, design, and procedures of the HTDS. On the subject of participation rates, Dr. Becker asked how many subjects were enrolled in the general information program on Hanford, such as newsletter distribution list or early mailing lists. Dr. Davis replied that probably there were few. The early mailing lists used were obtained from the HEDR project. Dr. Becker then asked if willingness to participate in HTDS was related to the amount of information that a person had obtained about Hanford. Dr. Davis replied that would be hard to determine.

On the subject of participation refusal rates, Dr. Schneider asked if there was anyone who said they already had thyroid cancer and did not want to participate. Dr. Kopecky said there was one. Dr. Davis commented there were more who said they did not have thyroid disease and therefore did not want to participate.

Dr. Becker commented that he thought it odd that a high proportion of study participants had CT scans and upper GI radiographs. Dr. Davis commented that the study team looked at data for some case-control studies being conducted in Washington and found similar proportions of participants in those studies having had CT scans and upper GI radiographs.

Dr. Hamilton then summarized definitions of disease outcomes used for the analysis. Each outcome definition had two components: the diagnosis based on examination and pathology information; and the basis for diagnosis, either exam, laboratory data, medical records, participant report. This approach provided information on presence or absence of disease, and 4 tiers of quality assessment in determining final disease outcomes.

Dr. Schneider commented that the definitions presented for benign thyroid nodules could be an area of criticism, especially a definition of benign nodule based on colloid abundance in the aspirate. In clinical practice, an FNA slide with abundant colloid and no cellular material should result in the patient proceeding to additional testing. Dr. Becker asked what quality assurance program was in place for cytopathology. Dr. Eheman asked if this would suggest a need for additional review of slides from FNA, and that there may be a need for additional discussion of quality assurance procedures in the revised final report.

Dr. Hamilton replied by summarizing the quality assurance plan for the cytology of FNA. The overriding objective had been that the study needed rapid response on the cytology and therefore used a single, experienced cytopathologist to review all FNA slides. Dr. Becker suggests an additional review is needed and suggested selecting a 5% sample of slides for additional pathology review. Dr. Hamilton replied that had been considered earlier and asked what

statistical purpose would that type of review serve. He thought that review might need 50%-75% of the slides to be re-read. Dr. Kopecky commented that a 5% sample would have identified only one of the twenty cancer cases. Dr. Becker replied that he did not see this as a confirming review, but rather a random check to assess the quality of the cytopathology.

Dr. Schneider asked what overlap was there in definitions for benign nodules and multi-nodular goiter.

Dr. Becker asked what definition of goiter was used. Dr. Hamilton replied it was two-fold enlargement of the thyroid gland without palpable nodules.

Dr. Needleman asked about how decisions were made on which nodules to aspirate in multi-nodular glands. Dr. Becker asked if most biopsies were done based on physical exam. Dr. Hamilton agreed, but commented about 15%-20% of FNAs were done based on ultrasound findings. He noted that less than 1% of the ultrasound determined abnormalities were greater than 1.5 cm. Dr. Schneider asked what would be results of analysis if the physical exam findings were ignored, and just the ultrasound exam results were used. He suggested separate analyses of physical exam and ultrasound exam findings should be considered.

Dr. Flanders asked what assessment had been done of reliability and agreement among physicians doing physical exams. Drs. Needleman asked about quality assurance for ultrasound technologists. How many ultrasonographers were used? What would be result if the ultrasonographer and examining physicians did not identify a nodule that the radiologist identified on later review? Dr. Hamilton replied that participants would have been called back for an FNA in that case.

Dr. Schneider commented there was a table shell listed in the analysis plan that was not included in the draft final report. That table shell was for the number of palpable nodules not confirmed by ultrasound.

Dr. Schneider suggested a diagram that summarizes the pathway to diagnosis for the various outcomes would helpful.

Dr. Hatch asked how feedback was given to ultrasonographers and physicians on quality assurance results.

Dr. Kopecky summarized the dose estimation procedures. Dr. Miller asked how the mean, median, and modal doses from each set of realizations was used.

Dr. Schneider asked for clarification on the 3 different ways to assign doses. Dr. Kopecky explained that the methods are determined by how much information is available. For example, if a CATI is not available for a participant, then there is no information on quantity of milk consumed, and the persons dose would be governed by residence and source of milk.

Dr. Devine asked if there was any pattern to residence near the boundary of HEDR domain for the out of area participants. Dr. Kopecky replied that a significant number lived close to the boundary in a few towns east and northeast of the domain.

Dr. Hatch asked about the proportion of study population for which they needed to use default values. Dr. Kopecky replied that was not in the report. He believed this did not have an impact on the dose response analysis, in contrast to the Utah thyroid cancer study. Dr. Davis replied that the dose schemes give a feel for the proportion with default values used**.**

Dr. Flanders asked what would be the effect if the dose response was done with mean dose rather than median dose. Dr. Kopecky replied that everyone would have a bit larger dose, so the dose response would be compressed. The summary text would not change.

Dr. Devine commented there should be more detailed discussion of multiple outcomes and the expected number of positive results.

Dr. Becker asked about data for non-thyroid outcomes. Dr. Kopecky replied there was only mortality data, although participants were asked in their interview if they had cancer. Dr. Becker pointed out that the local population was still very concerned about other outcomes. Dr. Kopecky noted there was an increase in deaths due to adverse birth outcomes, which could potentially be related to thyroid pathology.

Dr. Schneider asked about statistical significance of small ultrasound detected abnormalities. He stated it was not appropriate to summarize serum calcium as the only statistically significant analysis result. Dr. Kopecky disagreed, commenting that he believed the data presented were not inconsistent with noise. Dr. Becker stated he believed that follow-up of the participants with ultrasound detected abnormalities (UDAs) could be very important, since the clinical significance of these is unknown. Dr. Needleman commented that the public will only be satisfied when all leads are followed to dead-ends. Dr. Becker asked if some nodules or UDAs disappear. Dr. Needleman replied he did not believe any disappeared.

Dr. Eheman suggested doses be given for dichotomous exposure classification. She suggested that categorical analysis should be included in the report. Dr. Garbe noted that once the data files are made available for public use, there will be other investigators who will conduct these analyses. Drs. Davis and Kopecky replied they did not see how these could be helpful.

Dr. Devine asked why not use confidence interval instead of standard error. Dr. Hatch agreed, noting it gives a more readily comprehended assessment of variability in the effect estimate. Dr. Kopecky replied they could be done. Dr. Eheman stated there needed to be definitions of what p value is used to determine when something is statistically significant.

Dr. Schneider commented that the presentation on statistical significance of various results could be interpreted as trying to explain away any positive findings.

Presentation and discussion of the draft final report was completed at 12:30 pm. The peer reviewers convened separately to continue discussion during lunch.

Reviewer discussion

Dr. Hatch stated there were no glaring errors in the report.

Dr. Devine commented that he believed the discussion on uncertainty presented by Dr. Kopecky was sufficient. It should be summarized in writing with an appropriate graph for inclusion in the next revision of the draft final report. Dr. Devine noted he thought that the investigators would not need to include a summary of Nevada Test Site doses, given the uncertainty analysis presented this morning.

Dr. Hatch noted that mortality was 20% higher in the cohort than expected. The breakdown may raise questions from the public, in particular the mortality from congenital abnormalities. Dr. Becker noted his surprise at the high proportion of participants with certain thyroid conditions, such as autoimmune thyroiditis. This indicates a need for including some baseline data on thyroid diseases for comparison purposes. He suggested the data from Framingham, England, or NHANES could be useful for this.

Dr. Devine commented that the presentation of results on the SMR analysis will require a good deal of discussion and, potentially, some additional analysis.

Dr. Schneider noted that the interpretation of an acellular FNA biopsy as benign is not standard. A pathology review of FNA slides will be necessary, using 3 pathologists. He suggested reviewing FNA slides for all study participants with thyroid cancer and for a representative sample of all other participants who had FNA. He asked what information was available on the number of FNA slides determined to be inadequate.

Dr. Flanders commented there should be discussion of mis-classification of outcomes, specifically noting results of ultrasonography QC program on page 108 of procedures section.

Dr. Schneider suggested the report should include separate analyses of palpable nodules, UDA 1.5 cm or greater, and UDA less than 1.5 cm.

Dr. Devine repeated his earlier comment that the report needs more extensive discussion of multiple statistical comparisons. The report should note the number of significant findings that would be expected.

The group in general noted the public could perceive discussion of these results as an effort to explain away the findings. The reviewers recommended careful consideration of how the discussion of interpretation of negative findings in an epidemiologic study is framed.

Dr. Flanders questioned why logistic regression is adjusted for age when the dichotomous exposure analysis is done, and not otherwise.
Dr. Becker suggested CDC invite the public to sit down with us to review the results. He thought that should be relatively soon.

Dr. Schneider asked what was resolution on the Native American feasibility component. Dr. Garbe replied CDC had discussed that with the Study Management team the day before, and had reached agreement that we plan for a briefing to the ICHHP at an already scheduled meeting on December 10. [Note: This meeting with ICHHP was held on January 12, 1999].

The entire group re-convened at 2 pm to discuss reviewer comments.

Dr. Hatch asked Dr. Davis to comment on the 20% excess mortality in the cohort. Dr. Davis replied that the mortality analysis needed additional work. He commented that the last death certificates were not received until July 1998, which limited the amount of time for analysis. He welcomed any suggestions from reviewers on additional work or discussion of mortality data. Dr. Hatch asked if mortality analysis could be the subject of a second report. Dr. Garbe replied that additional mortality analyses should be done, but that the analysis in this report would need to be released along with the dose response analysis. Dr. Flanders asked what was the overall mortality in these counties. Ms. Adams Myers replied that most death certificates for the general population in these counties were not computerized.

Dr. Becker commented that overall he believed the study team had done a remarkable job and that they all should be commended. It is hard to find things to criticize in the report.

Summary comments

1. Uncertainty analysis presented today appears to be sufficient and needs to be written up for the release version of the report.

2. FNA cytopathology review is needed, using at least 3 cytopathologists. This should be a review of all persons with thyroid cancer, and a sample of all others who had FNA.

   Dr. Davis asked what would anyone do with this information. Dr. Hatch replied it gives an estimate of the error in the disease endpoint. Dr. Schneider noted this would allow a statement at release time that the cytopathology review is planned to assess if there were false negatives, that is persons who had FNA, but no surgery. Dr. Becker

commented that even thought the study team was confident of the pathology results, this review would provide confirmation, and it is easy to get. Dr. Davis replied he could see there might be perceptions of incomplete work. This is one area where they had not gone the extra mile.

3. There should be clearer explanation of why age adjustment was done in the dichotomous exposure analysis, and not elsewhere.

   Dr. Kopecky explained this was done because persons born in Stevens, Ferry, and Okanogan counties were only included in the Pilot Study. Therefore, their age at examination was younger than that for the remainder of the study population.

4. There were inconsistent applications of effect modification.

   Dr. Kopecky commented that this would be corrected.

5. Statistical models should be specified more clearly. It was difficult to tell when models were age-, sex-adjusted and when they were not.

6. There should be discussion of:

   a. Interpretation of the meaning of negative findings in an epidemiologic study;

   b. How findings that were observed will be followed-up;

   c. Multiple comparison issues;

   d. More discussion of the SMR results.

7. One strength of the study has been the openness of the process. There are findings reported here that people may not like and the report should be sure to emphasize areas that are re-assuring.

8. Greater care could be taken in the final version to be more consistent in the description of the statistically significant or borderline results.