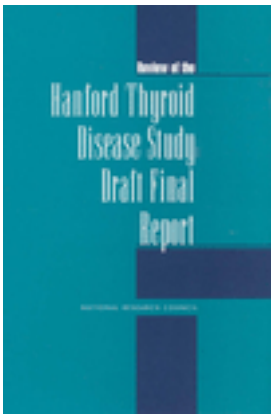

Review of the Hanford Thyroid Disease Study Draft Final Report

Committee on an Assessment of Centers for Disease Control and Prevention Radiation Studies from DOE Contractor Sites: Subcommittee to Review the Hanford Thyroid Disease Study Final Results and Report, National Academy of Sciences

ISBN: 0-309-57146-4, 228 pages, 6 x 9, (2000)

Review of the Hanford Thyroid Disease Study Draft Final Report

Committee on an Assessment of Centers for Disease Control and Prevention Radiation Studies from DOE Contractor Sites:
Subcommittee to Review the Hanford Thyroid Disease Study Final Results and Report
Board on Radiation Effects Research
Commission on Life Sciences
National Academy of Sciences

NATIONAL ACADEMY PRESS
Washington, DC 2000

NATIONAL ACADEMY PRESS 2101 CONSTITUTION AVENUE, NW WASHINGTON, DC 20418

NOTICE: The project that is the subject of this report was approved by the Governing Board of the National Research Council, whose members are drawn from the councils of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine. The members of the committee responsible for the report were chosen for their special competences and with regard to appropriate balance.

This report was prepared under the Centers for Disease Control and Prevention contract 200-95-0965 between the National Academy of Sciences and the Centers for Disease Control and Prevention.

International Standard Book Number 0-309-06883-5

A limited number of copies of this report are available from
National Research Council
Board on Radiation Effects Research
2101 Constitution Avenue, NW
Washington, DC 20418
(202) 334-2232

Printed in the United States of America

THE NATIONAL ACADEMIES

National Academy of Sciences
National Academy of Engineering
Institute of Medicine
National Research Council

The **National Academy of Sciences** is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. Upon the authority of the charter granted to it by the Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Bruce M. Alberts is president of the National Academy of Sciences.

The **National Academy of Engineering** was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Dr. Wm. A. Wulf is president of the National Academy of Engineering.

The **Institute of Medicine** was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, upon its own initiative, to identify issues of medical care, research, and education. Dr. Harvey V. Fineberg is president of the Institutedicine.

The **National Research Council** was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both Academies and the Institute of Medicine. Dr. Bruce M. Alberts and Dr. Wm. A. Wulf are chair and vice chair, respectively, of the National Research Council.

www.national-academies.org

COMMITTEE ON AN ASSESSMENT OF CENTERS FOR DISEASE CONTROL AND PREVENTION RADIATION STUDIES FROM DOE CONTRACTOR SITES: SUBCOMMITTEE TO REVIEW THE HANFORD THYROID DISEASE STUDY FINAL RESULTS AND REPORT

ROY E. SHORE *(Chair)*, New York University School of Medicine, New York, NY
ANDRE' BOUVILLE, National Cancer Institute, Bethesda, MD
BRUCE B. BOECKER, Lovelace Respiratory Research Institute, Albuquerque, NM
A. BERTRAND BRILL, Vanderbilt University Medical School, Nashville, TN
PATRICIA A.H. BUFFLER, University of California, Berkeley, CA
SHARON M. FRIEDMAN, Lehigh University, Bethlehem, PA
SUSAN E. LEDERER, Yale University School of Medicine, New Haven, CT
CARL M. MANSFIELD, University of Maryland Medical Systems, Baltimore, MD
DONALD E. MYERS, University of Arizona, Tucson, AZ
DANIEL O. STRAM, University of Southern California, Los Angeles, CA
ROBERT G. THOMAS, Folsom, CA

CONSULTANTS

HAROLD BECK, Environmental Measurements Laboratory, New York, NY, Retired
MAUREEN HATCH, Mt. Sinai School of Medicine, New York, NY
ARTHUR SCHNEIDER, University of Illinois College of Medicine, Chicago, IL
ERNEST MAZZAFERRI, Ohio State University, Columbus, OH, Retired

CLS ADVISER

BARABARA S. HULKA, University of North Carolina at Chapel Hill, NC

NATIONAL RESEARCH COUNCIL STAFF

STEVEN L. SIMON, Study Director
BRIDGET R. EDMONDS, Project Assistant
DORIS E. TAYLOR, Staff Assistant

SPONSOR'S PROJECT OFFICER

JAMES SMITH, Centers for Disease Control and Prevention

EDITOR

NORMAN GROSSBLATT

BOARD ON RADIATION EFFECTS RESEARCH

[**] Member of IOM
[*] New members effective 7/1/99

COMMISSION ON LIFE SCIENCES

NATIONAL RESEARCH COUNCIL STAFF

ACKNOWLEDGEMENTS

During the course of the committee's deliberations, several individuals provided information to the committee.

Appreciation for these contributions is extended to the following:

Harold Beck, Environmental Measurements Laboratory, New York, NY, Retired

Maureen Hatch, Mt. Sinai School of Medicine, New York, NY

Arthur Schneider, University of Illinois College of Medicine, Chicago, IL

Ernest Mazzaferri, Ohio State University, Columbus, OH, Retired

Owen Hoffman, Senes Oak Ridge, TN

William Farris, Pacific Northwest National Laboratory, Richland, WA

Trisha Pritikin, Berkeley, CA

Keith Baverstock, World Health Organization, Helsinki, Finland

Lynn Lyon, University of Utah, Salt Lake City, UT

Charles Grossman, Legacy Good Samaritan Hospital, Portland, Oregon, and Northwest Radiation Health Alliance

Robert Schenter, Pacific Northwest National Laboratory, Richland, WA

James Ruttenber, University of Colorado School of Medicine, Denver, CO

Tim Connor, North West Environmental Education Foundation, Spokane, WA

James Thomas, Short, Cressman and Burgess, Seattle, WA

Edward Liebow, Environmental Health and Social Policy Center, Seattle, WA

Larry Jecha, Benton-Franklin Health Department, Richland, WA

Judith Jurgi, Hanford Downwinders Coalition, Seattle, WA

Louise Kaplan, Pacific Lutheran University, Tacoma, WA

Bea Kelliegh, Hanford Health Information Network, Seattle, WA

Laura Chenet-Leonard, Oregon Health Division, Portland, OR

Elke Shaw-Tullock, Idaho Division of Health, Boise, ID

Lynn Stembridge, Hanford Education Action League, Spokane, WA

Kenneth Kopecky, Fred Hutchinson Cancer Research Center, Seattle, WA

Scott Davis, Fred Hutchinson Cancer Research Center, Seattle, WA

Karen Dom-Steele, Spokane Spokesman-Review, Spokane, WA

Annette Cary, Tri-Cities Herald, Spokane, WA

This report has been reviewed in draft form by individuals chosen for their diverse perspectives and technical expertise, in accordance with procedures approved by the National Research Council's Report Review Committee. The purpose of this independent review is to provide candid and critical comments that will assist the institution in making the published report as sound as possible

and to ensure that the report meets institutional standards for objectivity, evidence, and responsiveness to the study charge. The review comments and draft manuscript remain confidential to protect the integrity of the deliberative process. We wish to thank the following individuals for their participation in the review of this report:

John Buonaccorsi, University of Massachusetts, Amherst, MA
Darrell Fisher, Pacific Northwest National Laboratory, Richland, WA
Robert Griffin, Marquette University, Milwaukee, WI
David G. Hoel, Medical University of South Carolina, Charleston, SC
Henry D. Royal, Mallinckrodt Institute of Radiology, St. Louis, MO
Richard B. Setlow, Brookhaven National Laboratory, Upton, L.I., NY
Leonard A. Stefanski, North Carolina State University, Raleigh, NC
Steve Wing, University of North Carolina at Chapel Hill, NC

While the individuals listed above have provided constructive comments and suggestions, it must be emphasized that responsibility for the final content of this report rests entirely with the authoring committee and the institution.

The committee is very appreciative for the expertise, dedication, and hard work of the study director, Steven L. Simon. The attention to administrative details by Doris E. Taylor and Bridget R. Edmonds from the Research Council's Board on Radiation Effects Research is also appreciated. Sincere appreciation is also extended to Norman Grossblatt, who edited the report.

CONTENTS

Public Summary

BACKGROUND

In 1986, officials of the US Department of Energy revealed that the Hanford Atomic Products Operations in Richland, Washington, had been releasing radioactive material, in particular iodine-131, into the environment over a period of years. This information, which confirmed the suspicions of some people in the Pacific Northwest about what they called the Hanford Reservation or just Hanford, created quite a stir. Both the US Congress and citizens of the Northwest became keenly interested in knowing whether these radiation releases had caused human health effects. They were particularly concerned about whether Hanford releases of iodine-131 had led to an increase in thyroid disease among the population of the area.

In 1988, Congress ordered a study of the human health effects of exposure to the iodine-131 released from Hanford. Funded by the Centers for Disease Control and Prevention (CDC), the study was carried out by the Seattle-based Fred Hutchinson Cancer Research Center over the last decade. The study examined estimate of exposure[1] of the thyroid and rates of thyroid disease because iodine-131 concentrates in the thyroid and that organ would be the best indicator of radiation damage in the population.

[1] Although *dose* is the correct technical term, this summary will use *exposure* to refer loosely to a person's total radiation dose to the thyroid gland resulting from either short-or long-term exposure to iodine-131 in the atmosphere and environment from releases during the period 1944–1957.

Scientists have recognized for about 45 years that iodine-131 intake can lead to substantial radiation exposure of the thyroid and possibly to the development of thyroid cancer and other thyroid diseases. The likelihood that a given person will develop thyroid disease after being exposed depends on the size of exposure. The amount of radiation received by people living downwind of the Hanford site depended on specific characteristics of their individual lives, such as when they were born, where they lived, what foods they ate, and where they obtained those foods. The iodine-131 exposure of children occurred mainly through the milk they drank and to a lesser extent through the leafy vegetables and fish they ate. Breathing contaminated air also exposed Hanford area residents and was included in the exposure calculations. The radiation exposures of the thyroid glands of small children were, on the average, much higher than those of adults because children's thyroids are much smaller than those of adults and children consume a lot of milk.

To conduct the Hanford Thyroid Disease Study (HTDS), a 9-year $18 million effort, the investigators had to contact 5,199 eligible people who had been born near Hanford (in Franklin, Adams, Benton, Walla Walla, Okanogan, Stevens, and Ferry counties) in the period 1940–1946 because the period of greatest radiation releases was 1944–1947. Eventually, the HTDS investigators enrolled 3,441 subjects in the study, gave them extensive medical examinations to look for evidence of thyroid disease, and used a questionnaire on risk factors for thyroid disease. The HTDS investigators estimated individual radiation exposures for the 3,190 people who, during 1944–1957, had ever lived in the geographic area for which dose calculations were made. Estimating radiation exposures of 50 years ago is a daunting task for scientists because of the many unknowns about people's lives, habits, and diet. Nevertheless, a detailed method that had been developed previously by the Pacific Northwest National Laboratory was used to estimate the exposure received by each HTDS participant.

Armed with the rates of thyroid disease found among the 3,441 participants and estimates of radiation exposure received by 3,190 of them, the HTDS investigators used statistical methods to determine whether there was a relationship between the rates of disease found and the estimated radiation exposures. Ordinarily, one would expect that participants with larger radiation exposures would have higher rates of disease. The statistical analysis was complex for a number of reasons, including difficulty in determining the radiation exposure received by each person.

On January 28, 1999, the HTDS investigators and CDC released a Draft Final Report (FHCRC, 1999a) of the study to the public. The report was a draft because, although it had undergone internal review by CDC, it was still to be reviewed and subjected to scrutiny and comment by the National Academy of Sciences-National Research Council (NAS-NRC). The draft was released 2 months earlier than planned, for several reasons, including public pressure for the report's release without changes made by CDC and the desire by NAS-NRC to have an open review of the report. The primary finding of the HTDS draft report was that there was no evidence linking radiation exposure from Hanford to the rate of thyroid disease found in the study population. The lack of evidence of an effect, in scientific terms, is often called a "negative" finding. While presenting their findings to the media and regional citizen groups, the HTDS investigators overstated the certainty of their results.

Many Northwest citizens were upset not only about the findings of the study, but also about how the results of the study were conveyed by the investigators. Shortly after the draft's release, at CDC's request, NAS-NRC began an independent and comprehensive appraisal of the study methods, results, and interpretation and of how the study's findings were communicated to the public. This report is a fulfillment of that request.

The NRC subcommittee studied the HTDS Draft Final Report and discussed its contents in a series of meetings and e-mail communications over about 9 months, in February–October 1999.

The subcommittee arrived at consensus views on six specific questions asked by CDC: how well the HTDS investigators analyzed their data, how well the results were presented, whether their conclusions were reasonable, whether the material provided to the public was accurate and useful in helping the public to understand the study findings, how the presentation might need to be changed for the final report, and how CDC might improve communication with the public in the future. The subcommittee also developed a number of issues of its own to evaluate.

Detailed comments concerning the HTDS Draft Final Report are included in various chapters of the main report. The executive summary following this section highlights the views of the NRC subcommittee. Answers to the questions mentioned above are summarized in the executive summary and answered fully at the end of the subcommittee's report. This public summary is intended to review the main points of the executive summary in nontechnical language.

For its report, the NRC subcommittee concentrated on five main subjects for evaluation: design of the HTDS, estimated radiation exposures, data analysis, statistical power, and communication issues. Its major findings and recommendations appear below in boldface type.

DESIGN OF THE HTDS

The NRC subcommittee considered the HTDS design to be appropriate to address its goals. The methods to determine who the participants should be and where they were living were exceptionally good, and the HTDS collected the appropriate data on participants to enable the proper type of analysis. Although the subcommittee found the study methods to be of high quality, there are considerable uncertainties in some of the information.

The investigators chose the most relevant population to study: those in the most highly exposed areas who were young children at the time of the greatest iodine-131 releases. It was also

reasonable to study, as the subcommittee did, a low-exposure group upwind of and more distant from the Hanford site. The investigators were able to examine a high percentage of eligible persons, and this was a strength of the study. The information collected included such items as sex, history of other radiation exposures (such as from medical procedures), smoking history, and ethnicity.

Knowing the childhood milk-drinking habits of the participants in the study was particularly important because iodine-131 is most readily transferred to children through cow's milk as a result of the fallout that settles on pasture grass. The investigators attempted to question a parent or other close relative about each participant's residence history, where milk was obtained, and the amount of milk that was consumed during the period of the iodine-131 releases (1944–1957). If relatives were not available, then participants were given a questionnaire at the time of the medical examinations to get their history of residences and sources of milk. For 38% of the subjects, no parent or close relative was available to provide detailed information about childhood milk-drinking.

The NRC subcommittee found that the clinical examinations and laboratory studies were performed with good-quality, scientifically valid methods.

Ultrasound and palpation methods were used in the examinations. In palpation, a physician feels a person's thyroid gland in the neck with his or her fingers to determine its size and detect lumps. The subcommittee's only criticisms of the medical procedures were related to some quality-control procedures in the pathology review and to the fact that some requested medical records could not be obtained. But those criticisms were not important enough to invalidate the findings of the study.

Estimated Radiation Exposures

The NRC subcommittee's review found that the precision of the exposure estimates ranged from one-third or

one-half the best estimate to 2 or 3 times greater than the best estimate. That range is reasonable for historical-dose reconstructions. Evaluations of the model by other scientists have cast doubt on some of the factors involved in the model. The subcommittee also has concerns about some factors that might lead to greater overestimation or underestimation of the radiation exposures than was acknowledged by the HTDS investigators.

Pacific Northwest National Laboratory developed the computer model used to estimate the radiation exposures in the Hanford Environmental Dose Reconstruction (HEDR) project. This model had to take into account many factors: how much iodine-131 was vented from the Hanford site, the wind directions and other weather-related measures, how fast the iodine-131 settled to Earth, how much stayed on vegetation, how much vegetation was consumed by cows (which depended on the season), the fraction of the iodine-131 eaten or drunk by cows that was transferred to their milk, the length of time between when the farmer milked the cow and when the milk was consumed by a child, where the milk consumed by a child came from (for instance, a local versus a distant dairy), how much milk was consumed by the child at various ages, the fraction of the iodine-131 consumed (or breathed) that was deposited in the thyroid gland, and how long it stayed there. The model had to be able to estimate thyroid exposures of persons of different sexes, ages, places of residence, and dietary habits. The subcommittee found that the general method used in the model was suitable for the HTDS, assuming that the proper information about each participant could be obtained and used.

The NRC subcommittee found that the resulting exposure estimates for the HTDS participants were probably fairly accurate, mostly within a factor of 2 or 3. This statement is based on the results of validation exercises using the HEDR models (Napier and others, 1994). Recently, however, several scientists have claimed that the amount of iodine-131 released from the

Hanford site was higher than calculated by HEDR developers and that the HEDR model therefore underestimated the thyroid exposures by roughly 30%. And, the NRC subcommittee thinks that the model overestimates the iodine-131 that was transferred from pasture grass to cows' milk; this would mean that the model overestimated exposures. A careful reassessment of these elements of the model by the model developers is needed.

Errors like those can affect a study's findings about a relationship between disease rates and estimated radiation exposures. The ability of the HTDS to find the true relationship is called its "statistical power" and has been a focus of attention by the NRC subcommittee.

The NRC subcommittee found that the HEDR and HTDS investigators probably assessed individual exposures as being more precise than they actually were because some sources of uncertainty were underestimated or not dealt with.

The subcommittee noted that exposures that took place 40–50 years ago could not be precisely estimated and that such a situation could substantially reduce the ability of the study to detect a radiation effect. (Uncertainty and the power of the study are discussed further in this summary.)

The HTDS did examine the impact of fallout exposures from nuclear weapons tests conducted at the Nevada Test Site but overlooked the other sources of fallout exposure (such as nuclear tests in the Pacific and the Soviet Union). The NRC made a crude assessment of the exposures from global fallout and found that, on the average, the thyroid doses from global fallout were somewhat smaller than those from NTS fallout. In addition, the global fallout exposures occurred during the teenage years and early 20s among the study population. The NRC concluded that global fallout is not likely to have a large impact on the results of the epidemiologic study.

Analysis of HTDS Data

The subcommittee found some limitations in the HTDS data analysis, including exclusive use of the HEDR estimates of thyroid exposures from the Hanford releases, possible inaccuracies in exposure estimates for people who had lived only part of the time in the Hanford area, the need to analyze thyroid-disease rates by geographic area, and the absence of some key tables.

It is difficult to analyze the results of a study of the occurrence of disease if the number of cases is small. Although more than 3,000 people were evaluated for thyroid disease in the HTDS study, only 20 had thyroid cancer; and only 14 of those lived in the region covered by the HEDR model during 1944–1957 and could therefore have exposure estimated. The numbers were greater for most other thyroid diseases; for instance, benign thyroid nodules (noncancerous lumps) were found in 250 people. The radiation effect in causing this disease could be estimated with more certainty because of the larger number of cases.

The NRC subcommittee was critical of the HTDS investigators' exclusive use of the HEDR estimates of thyroid exposure for the data analysis and suggests supplemental analyses that could help to confirm or weaken the conclusions of the study. The subcommittee also found the analyses of the radiation effect (called "dose-response relationship" in the study) difficult to interpret for a variety of reasons. The subcommittee believes that a more complete analysis should be carried out to estimate exposures of people who were out of the study's geographic area for some of the time when the iodine-131 releases took place.

The subcommittee recommends that the HTDS investigators conduct more analyses to address the fact that the thyroid disease rates in the HTDS appeared to differ in unexpected ways between one geographic area and another. The geographic area in which each person was born should be taken into account to explain the unusual finding that thyroid disease rates tended to be

higher in areas that were expected to have the smallest amount of iodine-131 deposited on the ground.

The subcommittee believes that the HTDS investigators were correct in emphasizing analyses of the radiation effect rather than comparisons with another population. It does not believe that comparing the HTDS study group with some unexposed general population would be useful.

Members of the public have repeatedly questioned why no unexposed control group was involved in the HTDS so that disease rates could be compared. There are several major reasons why the panel does not think that that would be a valid comparison. First, for reasons unrelated to radiation, persons living in various geographic areas can vary in their likelihood of developing thyroid cancer. Second, the rates of disease found in the HTDS are based on thyroid examinations. Intensive medical examinations usually find more thyroid disease than would otherwise be known about from routine medical practice. Because no other population in the Northwest has been examined this way, a valid comparison with other populations cannot be made. Any conclusions drawn from comparisons with another population that is defined as a "control group" would have more potential for error than the conclusions drawn from the analyses that the HTDS investigators conducted. Third, the analysis of a radiation effect is a valid guide to the risk to the Hanford population even without the use of an unexposed control group, as long as there is a sufficient range of exposure levels and they are estimated with reasonable accuracy.

The subcommittee is concerned that the results of the study were reported—and interpreted—in black and white terms of whether a statistical test was passed or failed. It recommends that confidence limits be provided throughout the report to allow readers to judge how large a radiation effect might be consistent with the data. It feels that the HTDS

investigators probably overstated the strength of their finding that there was no radiation effect.

Usually, scientists provide both a best estimate and a range of estimates—called "confidence limits" or "confidence intervals"—to use to interpret statistical results. However, the HTDS investigators provided only their best estimate, not the confidence limits, for the size of possible radiation effects in the report or in their public statements. That made their findings seem more solid than they actually were.

Furthermore, the HTDS investigators should have calculated confidence limits that account for both the imprecision in the exposure estimates and the conventional statistical imprecision. By not presenting confidence limits, especially ones that consider imprecision in exposure estimates, the HTDS investigators overstated the strength of their main findings in the draft report.

Statistical Power and the HTDS Interpretation

The subcommittee believes that the assumptions used by the HTDS investigators to estimate the needed sample size and to calculate statistical power were incorrect; their assumptions did not acknowledge that exposures could be estimated only very imprecisely. The subcommittee found that HTDS ignored five sources of imprecision, which decreased the ability of the study to detect a small radiation effect. That means that the negative results that the study obtained are less definitive than the report and press releases stated.

Because the HTDS results found no increase in thyroid disease with an increase in radiation exposure from iodine-131, a critical issue is how to interpret those findings correctly. To evaluate the HTDS interpretation, the subcommittee asked a series of questions. For example, were the data good enough? Do the underlying patterns of exposure and disease agree or disagree with the negative findings? Was the statistical power of the study high enough to make the negative findings convincing? (The higher the

statistical power of the study, the more confidence people can put in the study's findings.)

The subcommittee reviewed the factors that influence statistical power, focusing on the impact of lack of precision in the thyroid exposures calculated by the HEDR project. It found that the statistical-power calculations made inadequate allowance for imprecision in the dose estimates. Given that situation, the subcommittee believes that the HTDS did not have as much statistical power to detect radiation effects as the investigators claimed. That means that the results of no effect ("negative" findings) reported by the HTDS are less definitive than the report and related public documents stated. Hence, this subcommittee recommends that, if possible, the HTDS investigators redo the statistical-power calculations to take into account all the sources of imprecision and that they reinterpret the study results in accordance with the limitations of statistical power.

The subcommittee believes that the findings of the HTDS cannot be reliably distinguished from the findings of the study of thyroid disease among children in Nevada and Utah who had been exposed to fallout resulting from atmospheric nuclear weapons tests conducted at the Nevada Test Site in the 1950s. A marginally positive radiation effect was found in that study. It is likely that, given the confidence limits for both studies, there would be an overlap, even though one appears positive and one negative. That is because the findings of both studies are very imprecise.

Communication of HTDS Results to the Public

The subcommittee believes that the original communication plan developed for the HTDS, particularly the parts that emphasized open public communication, was well developed and should have been moderately successful if implemented as planned. However, several factors led to an early release of a draft report, rather than a final report. When

the Draft Final Report was released, a number of communication errors were made that caused public outcry.

Compared with the history of less than open public information from the US Department of Energy and its predecessor agencies, the early plans by CDC and the HTDS investigators for open communication about the study were enlightened and promising. So were the decision to establish a citizen advisory group for the study and the apparent level of cooperation offered to various other citizen groups in the region over the years of the study. All those early efforts should have helped to build trust and credibility for the study.

Some of the public outcry on release of the draft report might have been avoided if the original communication plans outlined in the HTDS draft had been followed. The draft report outlined a good communication plan for its release, which included an admirable concern for translating the technical information in the report into an understandable booklet for the public and other efforts, including a Web site, to share information with the public. But the plan also called for delivery to the public of *final* information about the study, not a draft that had not been subject to review by outside scientists. Instead, several events forced the early release of the Draft Final Report and pre-empted the original communication plan.

Not only the early release of the report was a problem, but so was the main message in the report (namely, a strong statement that iodine-131 releases had caused no thyroid disease). In trying to decide how to present this message, CDC was on the horns of a dilemma. CDC personnel had been urged by some members of citizen groups not to alter the report before its release; they wanted the report to be released just as the HTDS investigators had written it. CDC also had to respect issues of academic freedom regarding the principal investigators' views. But after the draft's release, the CDC people were blamed for not intervening to counter the strong message delivered to the public by the HTDS investigators.

A key weakness of the communication effort surrounding the release of the draft report was that the public materials written and the oral statements made by HTDS investigators overstated the certainty of the study (the statistical power) and the conclusiveness of the negative findings, but did not report any of the uncertainties.

The public materials factually represented what appeared in the draft report. But, given the state of a draft document that had not been reviewed externally and a number of uncertainties in the data, the strong statements that the investigators made publicly were unwarranted. On the basis of comments received by the NRC subcommittee from members of the public, it is clear that many persons with an interest in the findings of the study were not only disappointed with the reported negative results, but also upset by how the results were disseminated and described.

A number of factors contributed to the problems surrounding the draft report's release, including (1) a perceived need for an information blackout that included the citizen groups that had been privy to most other parts of the study; (2) a complex schedule of briefings of groups in person in Washington, DC, and by telephone in the Hanford area to various state health agencies and citizen organizations only several hours before the media and public briefings on the findings; (3) a leak to the *New York Times* that related the findings to the public before most of the briefings in the Hanford area; and (4) a message that contradicted what most of the public thought would be the outcome of the study.

The subcommittee believes that in the media and public briefings the HTDS investigators paid insufficient attention to the audience's health concerns and fears and that HTDS investigators and CDC officials should have offered more balanced, and possibly alternative, interpretations of the findings and discussed their implications for individuals.

During the media briefing and public meeting held to announce the findings of the HTDS, the investigators emphasized

the overall statistical results of the study and did not seriously discuss the outcome for individuals. That approach angered many members of the community who had thyroid-related health problems. More care should have been taken to explain the differences between statistical relationships and individual outcomes. The subcommittee recommends that when the final report is released, implications for individuals and families that have suffered because of thyroid disease be explained and highlighted in the written materials and the public briefings. In addition, legitimate differences in viewpoints regarding study findings between the HTDS and CDC personnel should be explained and discussed.

The subcommittee recommends that a new communication plan be developed for the release of the final report, taking into account the serious problems encountered with the release of the draft report. In the final report and all public documents related to it, any important changes made from the draft report and all remaining uncertainties should be clearly outlined and explained. The subcommittee applauds CDC's open-communication policy and strongly recommends that this policy continue with the HTDS and similar studies.

The complicated briefing strategy used for releasing the Draft Final Report did not work well, and the subcommittee suggests that a more simple and efficient briefing plan be devised for releasing the final report. In particular, it recommends that telephone briefings be abandoned because all involved with release of the draft report disliked them. Citizen groups that have participated in a study over the years should not be kept out of the information flow concerning the study report's release until the very last minute, as they were with the briefings on the draft report.

The subcommittee also suggests that a small group of risk—communication experts, scientists, journalists, and citizens be convened to consider the more effective public release and discussion of controversial draft reports that have not been peer

reviewed, as well as other issues that could affect the future release of important CDC reports.

Concluding Remarks

In concluding its review of the HTDS Draft Final Report, the NRC panel considered the notion raised by the public that the HTDS is inconclusive in its findings. The subcommittee believes that the issue cannot be answered as simply as "agree" or "disagree", because the certainty of the interpretations from a complex study like the HTDS is always a matter of degree. The subcommittee members believe that the high certainty with which the HTDS investigators presented the negative findings of the draft report amounted to an overstatement. But the main finding of the HTDS final report could prove to be that no radiation effect can be observed. Given the imprecision in the exposure estimates and the effect of other statistical issues, the absence of any observable radiation effect is not proof that there is none. It does mean that the iodine-131 exposure did not have large effects. However, until estimates are given with appropriate confidence limits, we do not know how much risk to the thyroid is compatible with the data.

It seems doubtful that a better study could have been conducted in the downwind area, short of having some way to improve the exposure estimates greatly—an unlikely prospect because so little information is available on the exposures of 45 years ago. This carefully designed study, with sound followup and sound medical methods, has examined a large fraction of the most heavily exposed population and failed to find any obvious evidence of a radiation effect; that is, there was no evidence of abnormally high rates of thyroid disease in the Hanford "downwinders" examined who had the largest estimated exposures. Thus, at face value, the study was negative, and no increased risk was found. The pattern of individual exposures is in accord with such basic factors as the prevailing wind direction and distance from the Hanford site, and this accord generally supports the exposure modeling. Finding negative results of both

geographic and exposure comparisons implies that the iodine-131 exposures had no strong impact on thyroid disease.

However, if a similar exposure occurred elsewhere, one could not predict with confidence whether a positive or negative result would be seen. The small numbers of thyroid-cancer cases and the lack of precision in estimating individual exposures mean that one can have little confidence in the size of the risk estimates found in the HTDS.

At the time of the initial release of the Draft Final Report, it was indicated by the HTDS investigators that residents of downwind areas should feel relief that being close to the Hanford nuclear site did not result in increased risk of any thyroid disease. Such statements are too broad, but they might be reasonable in specific instances. For example, a healthy 55-year-old who lived near Hanford and drank a large amount of milk as a child can take comfort in learning that there is no evidence that he or she will have a greater risk of thyroid disease than other people in the general HTDS study area.

At various public-comment meetings, people who lived in downwind areas stated that their families experienced more thyroid disease than would have been expected in the population at large. Their disease could have been the result of unusual fallout or eating patterns or unusual susceptibility to radiation effects. But one should bear in mind that some cases of thyroid disease occur for reasons not understood by medical science. For example, thyroid disease tends to run in families, and family clusters could be related to genetic factors in the families or to chance. The lack of evidence of a dose-response relationship for any type of thyroid disease in the HTDS suggests, but does not prove, that the overall risk was not affected by Hanford fallout. The evidence does not rule out (although it does not support) the possibility that a weak association could affect, for instance, people who are already susceptible to thyroid disease because they are predisposed to it by genetic factors.

Executive Summary

BACKGROUND

It is well recognized that iodine-131, or ^{131}I (a radioactive form of iodine), is an important radionuclide because of the potential for human exposure to it after accidental releases from nuclear reactors and fuel—reprocessing plants. Deposition of ^{131}I onto pasture grass leads to contamination of cows' milk and ingestion of the radioactivity by humans. Because iodine is concentrated in the thyroid, radiation doses to the thyroid can result. The Draft Final Report of the Hanford Thyroid Disease Study (HTDS) describes a study of the cumulative incidence of thyroid disease and abnormalities among "downwinder" children exposed to ^{131}I from the Hanford Atomic Products Operations. Releases of ^{131}I began in December 1944 as a consequence of the chemical removal of plutonium from the fuel rods irradiated at the Hanford nuclear site.

The main study objective of the HTDS is described in the Draft Final Report as a "determination of whether thyroid morbidity is increased among persons exposed to releases of radioactive iodine from the Hanford nuclear site." In the study, 3,441 subjects who had been born near Hanford in 1940–1946 were contacted in the 1990s and taken to several locations for medical examination for thyroid disease. A detailed dose-reconstruction method developed by the Pacific Northwest National Laboratory was used to assign likely thyroid doses to study participants. Estimating dose also entailed querying parents about study participants' residence and milk-consumption history in 1944–1957, the period of ^{131}I releases.

The Centers for Disease Control and Prevention (CDC) asked the National Academy of Sciences-National Research Council (NAS-NRC) to give an independent appraisal of the study methodology, results, and interpretation and of the communication of the study results to the public. Specifically, it asked:

- Has the analysis been carried out appropriately and completely?
- Are the presentation and the discussion of results complete?
- Are the conclusions reasonable?
- Was the material accurate and appropriate in providing guidance to the public in understanding the study findings?
- If these messages about findings need to be amended, how should the revised messages best be communicated to the public?
- With regard to release of future study reports, how can CDC improve the public communication process?

This report constitutes the response of the NRC subcommittee to that request. To respond to the charge, the NRC subcommittee felt that it needed to go beyond the specific questions addressed to it by CDC and develop a broad understanding and critique of the HTDS and the Draft Final Report. As part of those activities, the subcommittee solicited comments from outside experts and members of the public primarily in a public meeting held in Spokane, Washington, in June 1999, where 14 scientists and members of the public made formal presentations to the subcommittee about various aspects of

the Draft Final Report. Other members of the public also spoke during four open-comment sessions at the meeting. In addition, efforts were made to evaluate all information materials prepared for the public and additional CDC communication plans. Information was gathered through interviews with journalists, members of concerned citizen groups in the Hanford region, members of the CDC scientific and media staff in Atlanta, and the HTDS investigators.

In this summary, the main points follow the structure of our report and are presented under several headings: epidemiologic and clinical methods and data collection, dosimetry, statistical analyses, statistical power and interpretation of the study, and communication of the study results to the public. We then provide a brief synopsis of our response to the questions raised by CDC.

EPIDEMIOLOGIC AND CLINICAL METHODS AND DATA COLLECTION

The HTDS eligibility criteria called for including all persons born in the early 1940s in the counties that were predicted to have the highest exposures to Hanford releases (Benton, Franklin, and Adams counties) and randomly selected subjects born in the same period in four counties that were expected to have intermediate exposure (Walla County) or low exposures (Ferry, Stevens, and Okanogan counties). Attempts were made to determine the vital status of all 5,199 eligible potential participants and, if they were alive, to trace and enroll them in the study. A total of 3,441 people received thyroid examinations; thyroid doses could be estimated for 3,190 of them.

The subcommittee considered the study design to be generally appropriate to address the aims of the study. The investigators also chose the best population to study, namely, those in the most highly exposed areas who were young children at the time of the greatest ^{131}I, releases. The low-exposure group, up wind and more distant from the Hanford site, was also a reasonable choice. However, the one significant weakness of the design was

the uncertainty in thyroid dose estimates, with resulting potential for exposure misclassification.

The epidemiologic methods were exceptionally good. The sample was based on an almost complete census of eligible subjects, and location and participation rates were high. There was a high level of quality control in the epidemiologic procedures, and interviews and clinical examinations were "blinded" (without knowledge of exposure or disease status). The HTDS collected data on an appropriate set of potential confounding variables (risk factors that might distort or mask findings), including sex; age at first exposure to ^{131}I from Hanford; age at examination; history of diagnostic, therapeutic, and occupational radiation exposures; smoking history; and ethnicity.

The questionnaire to elicit information from parents or surrogates on the participants' milk-consumption patterns was carefully designed and field-tested. However, there is substantial inherent unreliability in recall of dietary habits of 40–50 years ago, especially when someone other than the mother was interviewed, as was the case in 26% of the interviews. Compounding that problem is that for 38% of the participants no parent or surrogate was available to be interviewed.

Generally, the clinical examinations and laboratory studies were performed with good-quality methods. Subjects were given physical examinations, including thyroid palpation by thyroid specialists, ultrasonography, and appropriate thyroid-hormone and thyroid-antibody blood tests. The subcommittee has mostly minor criticisms of the clinical and laboratory procedures pertaining to inadequate quality-control procedures in assessing and reporting cytopathology results and the possibility that some past thyroid diagnoses were missed, inasmuch as it was impossible to obtain 7% of the death certificates and 37% of the requested historical medical records.

The HTDS investigators provided the subcommittee with additional tabulations for examining deaths in the group studied. Although there was a small increase in mortality, mostly

due to perinatal mortality and congenital anomalies, the tabulations indicate that the increase was not due to ^{131}I exposure, in that it occurred both before and during the time of the ^{131}I releases. However, a more detailed tabulation of perinatal deaths and congenital anomalies should be provided to help readers to interpret these data.

Conclusions:

- The study design was generally appropriate.
- An optimal study population was chosen: the most highly exposed young children.
- The epidemiologic methods were of excellent quality.
- The clinical and laboratory methods were appropriate and generally had good quality control.
- Some past thyroid diagnoses might have been missed because medical records and pathology slides were unobtainable.

Recommendations:

- An adequate review of the cytopathology results is needed.
- The HTDS investigators should indicate for how many potential past thyroid diagnoses they were unable to obtain any medical confirmation, with a breakdown by reported type of thyroid disease and dose.
- The mortality experience should be tabulated in more detail.

DOSIMETRY

The HTDS relied heavily on the Hanford Environmental Dose Reconstruction (HEDR) method, which in turn relied heavily on the use of environmental-transfer models. Models were necessary to estimate movement of ^{131}I in the environment because very few measurements of ^{131}I from the 1940s were available. The estimates of the thyroid doses for the 3,190 participants ranged from of 0.0008 mGy to 2,842 mGy, with a mean of 182 Mgy. The HEDR models have been subjected to numerous reviews and to independent testing. The NRC subcommittee found the dose assessment, on the whole, to be structurally sound for the estimation of thyroid doses, but minor errors have been found and doubts have been raised about the validity of some assumptions and of results for some environmental conditions. A review of some of the key parameter values that went into the HTDS dosimetric model showed that most of them were reasonable, and the resulting dose estimates are generally supported, at least to within a factor of 2 or 3, by the validation studies performed by the HEDR project.

Recently, several scientists have asserted that the source term (the amount of ^{131}I released by Hanford) was underestimated. However, even if those scientists' points are all valid, we estimate that it would increase the total ^{131}I releases during 1944–1957 by only about 30%. Systematic dose underestimation has implications for the statistical power of the study. If the doses were underestimated across the board, the study would have greater statistical power than was projected, in which case negative results of the study would be more persuasive. However, if there was variation by subject in the degree of dose underestimation, or in the degree to which subjects vary in sensitivity (because of age differences, and so on), these could result in reduced statistical power.

The subcommittee believes that the doses might have been overestimated, in that the HTDS used an estimate of the fraction of ^{131}I eaten by a cow that is transferred to milk that was

about twice as high as estimates in other studies. If the doses were overestimated, the statistical power of the study would be less than was stated in the HTDS Draft Final Report. Revisions of the model and recalculation of doses would be required to determine the net effect of those factors.

A critical feature of the dosimetry system is that the dose estimates have large uncertainties because they are based on mathematical models, not direct measurements. It is very likely that the uncertainties were underestimated by the HTDS because some sources of uncertainty were not taken into account. A notable deficiency was in accounting for the uncertainties in the sources and amount of milk consumption reported by parents or their surrogates. Those reports were 40–50 years after the fact, so one would expect appreciable unreliability in recall of milk-consumption patterns. The uncertainties would lessen the statistical power of the study and thereby make its results less definitive.

During the 1950s and early 1960s, the Hanford population was also exposed to ^{131}I from Nevada Test Site (NTS) fallout and global fallout from atomic and hydrogen-bomb tests over the Pacific, in the Soviet Union, and elsewhere. The HTDS team performed some analysis of the impact of NTS fallout, but not of global fallout. A rough assessment of the thyroid doses arising from global fallout was performed as part of our review. The subcommittee concludes that the variability in thyroid doses from those sources is much less than that in doses from Hanford fallout in the population being studied, so NTS and global fallout are probably not important confounders of the Hanford dose-response associations. Nevertheless, they should be examined carefully to be sure.

Conclusions:

- The dose-assessment methods and their implementation are difficult to review because the information is scattered among numerous documents.

- Dose estimates appear accurate to the degree one normally expects for environmental-dose reconstruction but minor errors in the parameter values used in the model need to be corrected.
- Dose uncertainties were underestimated because errors were not included for all the factors in the dosimetry model. This conclusion is the same as stated in a previous National Research Council (1999) letter report that reviewed the analysis plan for the HTDS.

Recommendations:

- A single document describing clearly the HEDR dose—assessment methodology, including uncertainties, and its implementation by HTDS should be prepared.
- Errors in the dose-estimation model should be corrected.
- All dose-related uncertainties should be taken into account.

STATISTICAL ANALYSES

The observed number of cases of thyroid cancer, the thyroid-disease end point of greatest interest, was very small: 20, of which only 14 had dose estimates. That makes it difficult to perform a meaningful statistical evaluation. For most other thyroid diseases, the numbers were more substantial, such as 250 cases of benign thyroid nodules.

The HTDS used the HEDR thyroid doses as its only, or at least primary, tool to describe patterns of likely exposure. The subcommittee believes that the HEDR dosimetry should not have been the sole method for evaluating the association between ^{131}I exposure and thyroid disease, and it suggests that supplemental

analyses be done that could help to confirm (or refute) the dose-response analyses. The dose-response results given are difficult to interpret; all one sees is dose-response regression coefficients with no tabulations to help in interpreting the factors that influence the results.

The subcommittee recommends that the following tables, which were absent from the Draft Final Report, be included in the revised report: frequency distribution of individual doses, observed and expected numbers of disease or abnormality outcomes in several dose categories, and average doses according to such important categories as year of birth and milk consumption in childhood. Other potential risk factors were evaluated as possible confounding or effect-modifying variables, but no tables were presented to show the results of those evaluations.

The HTDS investigators assigned thyroid doses only for periods when subjects lived as children in the geographic area for which exposures were estimated. They made no attempt to estimate out-of-area doses for persons who were out of the area for part of the exposure period and to perform sensitivity analyses to determine the impact of the missing doses. (See chapter 5.)

Given that thyroid-disease and thyroid-abnormality rates appeared to differ by geographic area, the subcommittee recommends alternative analyses to address the issue. A set of analyses stratifying on geographic area is needed because the HTDS investigators' tabulations showed that the rates tended to be higher in areas with low fallout, so the geographic variations due to factors other than dose would induce a negative association between ^{131}I and thyroid-disease rates. That feature of the data could explain why most of the dose-response estimates were in the negative direction; the question is whether removing its influence would yield a positive association.

A concern expressed by members of the public is that a study in which everyone is exposed is not valid and that a completely unexposed "control" group is needed for proper assessment of the risk associated with Hanford ^{131}I fallout.

However, for several reasons, the subcommittee believes that the HTDS investigators were correct in emphasizing dose-response comparisons in the study rather than comparisons with other general populations. The slope of the dose-response curve can provide a valid index of the risk even without an unexposed control group, as long as a sufficient range of doses are estimated with reasonable accuracy. Comparisons with an external, general population are problematic on several accounts. Persons living in various geographic areas often vary in their baseline risk of thyroid diseases because of differences in dietary iodine intake and other unknown factors. The rates of detected disease in the HTDS are based on thyroid examinations and depend on the methods and criteria of those examinations. That often produces a large screening effect (detection of cases of disease that otherwise would not have been detected until some years later, if at all), so comparisons with rates from other geographic regions without comparable screening are not valid. Conclusions drawn from comparisons with general-population prevalence have more potential for bias than those drawn from dose-response comparisons in the study population.

The HTDS investigators analyzed the effect of interview versus default data for milk consumption, although results were not presented. The subcommittee suggests that they also examine associations, using only those with interview information to minimize dose misclassification. Another suggested alternative is to conduct dose-response analyses that stratify on interview versus default milk values.

Tabulation of thyroid-disease rates by reported milk-drinking habits is suggested to elucidate further whether ^{131}I exposure estimates tend to coincide with disease. The fact that higher thyroid morbidity was found in the less-exposed counties argues that the pattern of thyroid morbidity did not tend to track the likely geographic pattern of exposure to ^{131}I. However, birth location was not as important a determiner of thyroid dose as were

behavioral habits (as can be seen by the large range of doses in all counties), primarily milk-drinking.

Besides the lack of some important tables in the analysis, the subcommittee has concerns about the manner of presentation of the study results. The results were reported in black and white terms: a statistical test did or did not reject the null hypothesis. No confidence intervals on the estimates of the size of effects were given. The subcommittee recommends that confidence levels be provided throughout the report. Furthermore, had the investigators presented confidence intervals for the dose-response model on the basis of their statistics, the confidence intervals would almost certainly have been too small because dosimetry errors were ignored. A sophisticated statistical method described in the statistical chapter of the report would have at least partly taken dose uncertainties into account in the confidence intervals, but it was not implemented.

Conclusions:

- The HTDS report relies too heavily on dose-response analyses without providing sufficient associated evidence from tabulations of factors that could illuminate the results obtained. (See chapter 5.)
- A number of key tables were absent from the report, for example, tables of frequency distribution of doses; of observed and expected frequencies of each thyroid disease by, say, quartiles of dose; of thyroid disease rates by milk-drinking habits and other risk factors in disease; and of average doses by year of birth, amount of milk consumption in childhood, and the like.
- Differences in thyroid-disease rates by geographic area (called "geostrata" in the Draft Final Report) might have been an important confounder of the dose-response association.

- The subcommittee believes that the HTDS investigators rightly chose to emphasize comparisons of thyroid-disease rates in the study group rather than comparisons with external general population rates, because there is a potential for serious biases in the latter due to geographic variations and, especially, screening effects.
- Too little attention was paid to the range of risk estimates that is compatible with the data. Confidence intervals were seldom given.

Recommendations:

- A number of key tables should be included in the final report to help readers to interpret the dose-response results.
- The HTDS investigators should report on those who were out of the dosimetry area for part of the exposure period and examine the impact of the assumption of zero dose received during such periods.
- Analyses designed to control for possible confounding by geographic area should be conducted.
- There should be a more thorough set of analyses of thyroid-disease rates by milk-drinking information.
- Confidence intervals should be given and used in the interpretation of the study results.
- The confidence intervals should take into account all the sources of uncertainty in the dose estimates.

STATISTICAL POWER AND INTERPRETATION OF THE STUDY

The HTDS investigators were generally successful in achieving the sample size and dose distribution that they had projected as necessary if the study were to have adequate statistical power. The Draft Final Report indicates that for many of or all the thyroid end points the estimated power of the study was good (over 90%) for detecting plausible nonzero dose-response relationships. Nevertheless, the HTDS investigators' discussion of statistical power did not present how small the expected excess of thyroid cancers was. Based on their assumptions about the risk coefficient, the dose distribution, the number of persons in the study and the length of the follow-up, about 34 thyroid cancers were expected in the study, of which 19 would have been expected without any ^{131}I exposure and 15 were due to radiation exposure. Had these numbers been presented, they might have tempered CDC's evaluation of the scientific value of the study in relation to its cost. Our subcommittee reviewed the factors that influence power, focusing on the impact of uncertainties in the HEDR thyroid doses. We note that neither the power calculations nor analytic techniques used made explicit allowance for uncertainties in the dose estimates.

The investigators ignored five sources of uncertainty. First, in making their projections they assumed that the dose-measurement error was all of a type ("Berkson error") that would not reduce the strength of associations. That is different from classical measurement error, which does weaken the strength of associations and therefore requires a larger sample or a wider dose range to attain adequate statistical power. Individual-based dose error—for example, the uncertainty of a subject's milk-drinking habits—represents classical measurement error that needs to be taken into account in estimating the statistical power of a study. Second, multiplicative (as opposed to additive) dose uncertainties (such as errors in the source term or in the coefficient of transfer from cow intake to milk) that apply to everyone will add error to risk estimates and thereby decrease statistical power. Third,

correlations among multiplicative dose uncertainties can further reduce statistical power. Fourth, geographic variations in the baseline rates of disease can decrease statistical power if not controlled for. Fifth, some sources of uncertainty in the ^{131}I environmental pathway apparently were not included in the dosimetry uncertainty estimates.

Because those types of uncertainty were not taken into account in the statistical-power calculations, the subcommittee believes that the HTDS projections of statistical power are overestimated, perhaps substantially. The negative results of the study are therefore less definitive than the Draft Final Report and press releases stated.

The uncertainties listed above also have an impact on the width of confidence intervals around the estimates of thyroid-disease effects, so the study is less clearly negative than was portrayed. The confidence intervals would be more compatible with (although the best estimate does not support) the larger risks seen in other studies, such as the Utah NTS fallout study and the large pooled study of thyroid-cancer risk associated with external radiation exposure.

Conclusions:

- The HTDS investigators were successful in achieving the sample size and dose distribution that they projected as necessary if the study were to have adequate statistical power.
- However, the HTDS assumptions regarding statistical power did not include the possibility that dose uncertainty would weaken the associations. They ignored several sources of uncertainty that probably decrease the statistical power of the study.
- The subcommittee believes that the HTDS projections of statistical power are overestimated, perhaps substantially.

The negative results of the study are therefore less definitive than the Draft Final Report and press releases stated.

- The HTDS results are probably compatible with the risk estimates from the Utah NTS fallout study, because the various uncertainties would yield wide confidence intervals.

Recommendations:

- The HTDS investigators should describe the sources of uncertainty in as quantitative terms as possible and interpret their results in the light of these uncertainties.
- The HTDS investigators should recalculate the statistical power of the study, taking into account the dose uncertainties if this proves feasible.
- The compatibility of the HTDS study with other studies of radiation and thyroid disease should be re-examined, taking into account the impact of dose uncertainties.

COMMUNICATION OF THE STUDY RESULTS TO THE PUBLIC

Compared with the history of a less-than-open public-information policy of the Department of Energy and its predecessor agencies, the early plans by CDC and the HTDS investigators for open communication about the study were enlightened and promising. So was the decision to establish a citizen advisory group for the study and the apparent level of cooperation offered to various other citizen groups in the region over the years of the study. All those early efforts seemed to build trust and credibility for the study.

However, when the Draft Final Report was released, a number of communication errors were made that caused a public outcry. The draft report outlined a good communication plan, which included an admirable concern for translating the technical information in the report into an understandable booklet for the

public, and included a Web site to share information with the public. But several events forced an early release of the Draft Final Report, which pre-empted the original communication plan.

The main message of the report was problematic. The written materials and oral presentations made by HTDS investigators overstated the certainty (the statistical power) of the study and the conclusiveness of the negative findings. Although the public materials factually represented what appeared in the Draft Final Report, the strong statements made publicly were not tempered by expressing the uncertainties.

In trying to decide how to present the study, CDC was on the horns of a dilemma: some members of citizen groups had urged agency personnel in advance not to alter the report before its release, and CDC had to respect issues of academic freedom regarding the investigators' views; but after the report's release, they were blamed for not intervening to counter the strong message delivered to the public by the HTDS investigators.

On the basis of comments received by the subcommittee from members of the public, it is clear that some people with an interest in the findings of the study were disappointed with the reported (negative) results and upset by how the results were disseminated and described. A number of factors contributed to and complicated the problems surrounding the report's release: an information blackout that included the citizen groups, a complex briefing schedule by telephone in the Hanford area to various state health agencies and citizen organizations only several hours before the media and public briefings on the findings, a leak to the *New York Times* that related the findings to the public before most of the briefings in the Hanford area, and a message that contradicted what most of the public thought would be the outcome of the study.

Because serious problems were encountered in the scheduling and conduct of the prerelease briefings, a different briefing strategy should be used in the future, and telephone

briefings should probably be abandoned because they were disliked by all involved.

Delivering an unpopular message requires sensitivity to the audience's health concerns and fears. But the media and public briefings, and all written materials emphasized the overall statistical results of the study and did not seriously discuss the outcome for individuals. The implications for individuals and families that have suffered because of thyroid disease could have been explained in the written materials and public briefings.

The subcommittee applauds CDC's open-communication policy and strongly recommends that this policy continue with the HTDS and similar studies. It recommends that a new communication plan be devised for the release of the final HTDS report, taking into account the problems that have already been encountered. The final report should outline and explain any significant changes made in the Draft Final Report.

The subcommittee suggests that CDC convene a workshop of risk-communication experts, scientists, journalists, and citizens to discuss how to publicly release and discuss controversial unreviewed draft reports more effectively and to discuss other issues that could affect the future release of important CDC reports.

Conclusions:

- The early enlightened plans by CDC and the HTDS investigators for open communication about the study and for a citizen advisory group for the study should have helped to build trust and credibility.
- Early release of the Draft Final Report and public concerns about CDC changes in that draft led to many of the communication problems that resulted from the draft report's release.

- The information blackout and complex briefing schedule for release of the Draft Final Report worked against effective communication of the report's findings to the public and unnecessarily upset Hanford-area citizen groups that had cooperated with the HTDS over the years.
- A key weakness of the communication effort surrounding the release of the Draft Final Report was that the report and the public communications by HTDS investigators overstated the certainty (the statistical power) of the study and the conclusiveness of the negative findings and failed to discuss the uncertainties. CDC officials should have expressed their own interpretations—in addition to those of the HTDS investigators—about the draft report in the briefings and public documents.

Recommendations:

- Delivering an unpopular message requires sensitivity to the audience's health concerns and fears. In communications about the HTDS final report, implications for individuals and families that have suffered because of thyroid disease should be carefully explained. If there are plausible alternative interpretations of the results, they should be acknowledged.
- The subcommittee supports CDC's open-communication policy and strongly recommends that it continue. It recommends that a new communication plan be devised for the release of the final HTDS report and accompanying public documents, taking into account the problems that have already been encountered.
- In the HTDS final report and all public documents, any significant changes made from the Draft Final Report should be clearly outlined and explained, and all remaining uncertaintites should be noted and explained.

- Careful consideration should be given to how to release controversial reports to the public more effectively. The subcommittee suggests that CDC convene a workshop to discuss this and other communication issues of concern.

SUMMARY OF RESPONSES TO THE CDC'S QUESTIONS

The subcommittee's responses to CDC's six questions are summarized below.

Question 1. Has the analysis been carried out appropriately and completely?

Our overall assessment is that the epiderniologic and clinical components of the study were of excellent quality, including the study design, followup success, subject participation rate, interviewing, thyroid examination, and laboratory methods. The design of the dose-assessment model has been found, on the whole, to be reasonably sound for the estimation of thyroid doses, but several questionable assumptions have been identified that would have some impact on the estimated individual doses. The estimated-dose uncertainties that the HEDR project produced and the HTDS study used are underestimates of the total dose uncertainty because some significant sources of uncertainty were overlooked.

The basic objective in the statistical analysis was to determine whether there was an association between the occurrence of various thyroid diseases and exposure to ^{131}I released from Hanford. That was appropriately addressed by modeling the relationship between the frequency of a thyroid disease and dose, with consideration of an appropriate set of potential confounding variables.

Several other analyses that were not presented could have aided in the interpretation of the apparently negative results, a

critical one being an analysis of thyroid dose and disease with apparent geographic variations in disease rates controlled for.

The investigators made no attempt to model the out-of-area doses for persons who were included in the main analyses. Their approach to that issue could have led to attenuated results in that it potentially estimated the total fallout doses for some people but only partial doses for others. The impact of global fallout on variations in thyroid-disease risk should also be analyzed.

The subcommittee believes that the HTDS emphasis on analyses of subjects in the study rather than on comparisons with the general population is appropriate, inasmuch as the latter are potentially subject to serious biases.

Question 2. Are the presentation and the discussion of results complete?

One serious gap is that the methods used to calculate doses and uncertainties are not clearly or fully described in the dosimetry documents provided to the subcommittee.

A number of additional tables are needed. A tabular presentation of the pathways to diagnosis would help readers to assess how the final diagnoses were assigned. A table of the frequency distribution of doses would be informative. Similarly, tables that show observed and expected numbers of disease outcomes according to four or five dose groups would normally be expected. A description of the estimated dose distribution according to such important categories as geostratum, year of birth, and amount of milk consumption in childhood would be helpful.

The discussion of the results was substantially incomplete in that little was said about whether the confidence intervals were wide enough to be compatible with other parallel studies. Most important, there was no adequate discussion of how dosimetric uncertainties might have affected the confidence intervals and the statistical power of the study.

Question 3. Are the conclusions reasonable?

In concluding its review of the HTDS Draft Final Report, the subcommittee considered the notion raised by the public that the HTDS was inconclusive in its findings. The subcommittee believes that the certainty of the interpretations from a complex study, such as the HTDS, is always a matter of degree. Its members believe that the high certainty with which the HTDS investigators presented the negative findings of the draft report was overoptimistic. Still, the main finding of the final HTDS report might indeed be that no radiation effect could be observed: the lack of evidence of a dose-response relationship for any type of morbidity suggests that overall risks were unaffected by Hanford releases. Given the substantial degree of imprecision in the exposure estimates and the effect of other statistical issues, the absence of any observable radiation effect does not rule out the possibility that a small effect exists, although it does mean that large effects of the ^{131}I exposure can be excluded as incompatible with the data. Until estimates are given with appropriate confidence limits, we will not know how much risk to the thyroid is compatible with the data. The evidence does not rule out (although it provides no particular support for) the possibility of a weak association that could affect, for example, those already susceptible to thyroid disease because of predisposing genetic factors.

This carefully designed study, with sound followup and medical methods, has examined a substantial fraction of the most highly exposed population and failed to find any obvious evidence of a radiation effect; that is, there was no evidence of abnormally high rates of thyroid disease in the Hanford "downwinders" examined who had the largest estimated exposures. Thus, at face value, the study was negative and found no increased risk. The pattern of individual exposure estimates is in accord with such basic factors as the prevailing wind direction and distance from the Hanford site. Finding negative results of both geographic and

exposure comparisons implies that ^{131}I had no strong impact on thyroid disease. If a similar exposure occurred elsewhere, one could not predict the results with confidence. The small numbers of thyroid-cancer cases and the lack of precision in estimating individual exposures mean that one can have little confidence in the risk estimates found in the HTDS. At the various public-comment meetings, people who lived in down-wind areas stated that they and their families experienced more thyroid disease than would have been expected in the population at large. That could be due to genetic factors in families or even to chance, but the possibility that their disease was the result of unusual fallout or ingestion patterns or of unusual susceptibility to a thyroid radiation effect cannot be excluded.

Question 4. Was the material accurate and appropriate in providing guidance to the public in understanding the study findings?

For the most part, the written and oral messages about the Draft Final Report were accurate, but they were occasionally misleading in that they included statements that were too strongly worded, given the uncertainties that applied to the study. Keeping the study process and activities as "transparent"—that is, open visible and understandable—as possible for the public is a valuable approach that should not be abandoned because of the problems encountered with the release of the Draft Final Report.

Question 5. If these messages need to be amended, how should the revised messages best be communicated to the public?

Given all the communication problems that resulted from the release of the Draft Final Report, another detailed communication plan needs to be drawn up for release of the final report, including planning for unanticipated situations. Messages Must take into account the various audiences being addressed and

show concern and sensitivity for the thyroid-health issues that peopleperceive affect them. The full picture of the study results should be given, including all the uncertainties and other problems. A plan to brief the active citizen groups should be developed so that they have enough information to be able to respond to media inquiries about the report. In addition, an embargoed release of the report to journalists should be used so that they have a few days to read through the report and develop informed questions before the briefing.

Question 6. With regard to release of future study reports, how can CDC improve the public communication process?

The briefing structure should be simplified to try to eliminate leaks, and citizens who have participated in the advisory process all along should be given higher priority in the briefing structure.

It could prove helpful to CDC to conduct a workshop on improving the public-communication process that includes experts in risk communication, journalists, outside scientists, and members of citizen groups.

1

Introduction

The Centers for Disease Control and Prevention (CDC) sought an independent evaluation of its Hanford Thyroid Disease Study (HTDS) by the National Academy of Sciences National Research Council. In January 1999, at the time of the release of the Draft Final Report of that study, it addressed three questions to the Research Council's Committee on CDC Radiation Studies:

Has the analysis been carried out appropriately and completely?

Are the presentation and the discussion of results complete?

Are the conclusions reasonable?

In April 1999, CDC, prompted by its own concerns and those of interested members of the public, asked the committee to review and comment on material that had been prepared by CDC and the Fred Hutchinson Cancer Research Center (FHCRC) and provided to the public at the time of release of the HTDS Draft Final Report or thereafter. The material included a congressional briefing, overhead slides, a CDC press release, HTDS fact sheets, a series of CDC fact sheets, and a mailed update on the HTDS Draft

Final Report. The committee was asked to consider three additional questions:

Was the material accurate and appropriate in providing guidance to the public in understanding the study findings?

If these messages about findings need to be amended, how should the revised messages best be communicated to the public?

With regard to release of future study reports, how can CDC improve the public communication process?

2

Overview of the HTDS Draft Final Report and Organization of the Committee's Report

The HTDS was conducted as an epiderniologic followup study of the prevalence[1] of thyroid disease among those born in 1940–1946 in seven counties in the state of Washington. The counties were chosen for the likelihood of having residents who received high (Franklin, Adams, and Benton counties), intermediate (Walla Walla county), or low (Okanogan, Stevens, and Ferry counties) radiation doses to their thyroids from iodine-131 (^{131}I) released from the Hanford facilities.

The research was conducted in two phases: a pilot study that was completed in 1994 and the full study, which is the subject of the Draft Final Report.

It is well recognized that ^{131}I is particularly important with respect to human exposure to radionuclides. That is because of the existence of the pasture-cow-milk-thyroid pathway: ^{131}I deposited on grass can be eaten by cows, be secreted into the cows' milk, be consumed by people, and result in substantial doses to the thyroid as it is efficiently taken up by this gland. On the basis of data from the atomic-bomb survivors and other studies of radiation exposure, as described in the background section of the HTDS Draft Final Report, young children are considered to be more sensitive to thyroid disease as a consequence of exposure to

[1] The NRC committee uses the term "*prevalence*" loosely as a convenient way to refer to "cumulative incidence", which is what was actually assessed by the HTDS.

radiation than are older children or adults. Therefore, the HTDS focused on people who were young children at the time of the releases. The HTDS Draft Final Report comprises 10 sections and an executive summary, references, and several appendixes. Section I introduces the HTDS, which began in 1989 and was conducted by investigators at the FHCRC and the University of Washington under contract with CDC.

Section II provides background information on the Hanford nuclear site and activities that led to the establishment of this epidemiologic study. Of particular importance was the Hanford Environmental Dose Reconstruction (HEDR) project, which was started in 1987 to develop estimates of radiation doses that people might have received from Hanford operations. Radiation doses to the thyroid from ^{131}I have been the main ones. The HEDR results were critical for the dose-response analyses conducted in the HTDS. Section II also includes descriptions of the various thyroid diseases and other conditions that were studied as possible outcomes of irradiation from internally deposited ^{131}I.

Section III discusses several objectives of the HTDS. The primary objective was to determine whether thyroid morbidity was increased among persons exposed to ^{131}I released from the Hanford nuclear site in 1944–1957.

Section IV provides information on the study design. It discusses why the eligibility criteria were related to place and time of birth. An evaluable participant was defined as one who could be located and for whom sufficient information on thyroid disease and for determining radiation dose could be obtained. Outcome criteria, which define diagnostic criteria for various thyroid and parathyroid diseases and other changes, are also provided.

Section V summarizes the field procedures and methods and the results of data collection, including detailed information on how the cohort was defined and its members were identified, how study subjects were traced and recruited, how telephone interviews were conducted, and how doses were estimated. Attempts were made to determine vital status and to

trace and contact all the living among the 5,199 eligible potential participants. The section also provides information on scheduling, clinical investigations, interviews, medical reviews, determination of final diagnoses, and management of medical records.

Section VI discusses three special considerations related to the conduct of the HTDS. The first was an assessment of the feasibility of conducting a similar health study in the nine American Indian tribes and nations near the Hanford site; it discusses the steps taken to determine the feasibility and the decision that such a study would have insufficient statistical power to detect an increase in thyroid disease caused by Hanford releases. The second was a CDC-appointed advisory committee; it discusses the role of this committee in the design and conduct of the study, the schedule and locations of committee meetings, and the openness of the meetings to the public. The third was provision of information to the public throughout the whole HTDS process; it describes the approaches used—such as newsletters, fact sheets, and a telephone line—to keep the public informed of the activities and results of the HTDS.

Section VII describes the statistical methods used in the HTDS analyses. The information provided is related generally to the tests of the statistical significance of exposure-response relationships for various thyroid diseases, including an examination of possible confounding or effect-modifying factors. The data collected were in three categories: process information, characteristics of living evaluable participants, and analyses of exposures and outcomes. The analytic methods used to summarize the data are described in detail, as are the calculations made to examine uncertainties in dose. The possible confounding or effect-modifying factors included sex, age at first exposure to ^{131} I, ethnicity, smoking, and other radiation exposures. With respect to the last factor, exposure to ^{131}I, in fallout from weapons tests conducted at the Nevada Test Site (NTS) is given particular attention.

Section VIII presents the results of the HTDS. Given first are the characteristics of the 3,441 living evaluable study participants, including year of birth, age at examination, race or ethnic origin, medical radiation exposure, occupational history, and smoking history. Radiation doses to the thyroid from Hanford ^{131}I are summarized on the basis of calculations derived with the CIDER computer program developed as part of the HEDR project; these calculations were based on a person living "in area" or "out of area" from December 1944 to the end of 1957. For each study participant, 100 dose estimates were calculated, and the median of the 100 estimates was used as the best estimate of the person's dose. The implications of the number of persons studied and their calculated thyroid doses relative to the statistical power of the HTDS results are discussed. These dose values were used with a large number of outcome variables to conduct dose-response analyses. The outcome variables consisted of 11 categories of thyroid disease, ultrasonographically detected abnormalities, hyperparathyroidism, and various thyroid-related laboratory tests. Definitions are provided for each outcome variable, as are the results of the dose-response analyses. Some of the diagnoses were rare (for example, there were only 20 thyroid cancers); others were common. For none was the dose-response trend statistically significantly positive; and for several, the estimate of the linear slope was negative. In addition to the basic analyses, results of alternative dose-response analyses and other factors are presented. Information is also given on the patterns of mortality in the HTDS cohort.

Section IX discusses the results of the HTDS. It first summarizes the study accomplishments, including the identification and location of members of the cohort. Other aspects discussed include telephone interviews, medical evaluations of the cohort, and the successful location of a large proportion of the related medical records. Results of the dose-response analyses are summarized: there was no relationship between thyroid radiation dose from Hanford and the cumulative incidence of any of the 13 primary outcomes even when alternative analytic approaches were

used. The section discusses the possible influences exerted by such factors as the definition and selection of the cohort, the definition or misclassification of outcomes, the estimation of thyroid radiation dose, and uncertainty. It compares HTDS results with findings in other populations that were subject to irradiation of the thyroid, including the Japanese atomic-bomb survivors, Marshall Island residents exposed to fallout from the Castle-Bravo test, residents of Utah exposed to ^{131}I from atmospheric releases from the NTS, and people who lived near the Chernobyl nuclear reactor at the time of its catastrophic release of radionuclides.

Section X discusses communication of the HTDS results with the public. It summarizes ways that the HTDS staff used to maintain open and frequent communication, including public meetings, presentations at scientific meetings, interviews, fact sheets, and a toll-free 800 number. The remainder of this section is devoted to plans for communicating the study results to five targeted groups: study participants, the public and the mass media, the scientific community, the regional medical community, and government officials and agencies.

To provide a thorough and balanced review of the Draft Final Report and its communication to the public, the subcommittee undertook a number of activities. We met on February 4–5, 1999, in Atlanta, Georgia, on March 29–30, 1999, in Augusta, Georgia, and on August 30–31, 1999 in Washington, DC, to review the report; during the same period, we requested additional information from the HTDS and HEDR investigators.

We met in Spokane, Washington, on June 18–19, 1999; at an all-day open public meeting on June 19, attended by about 60 people. We heard from various experts and members of the public who wanted to present information regarding the HTDS Draft Final Report. During the day, 14 experts and interested members of the public invited by the subcommittee—some at the suggestion of citizen action groups—presented their views in person and through conference calls. Many of them also provided written statements. In addition, four public-comment sessions allowed any member of

the public to have his or her views heard on a variety of subjects in the HTDS Draft Final Report. (Appendix A contains the agenda and a list of speakers.)

To gather additional information for the communication section of our report, we conducted telephone interviews with two journalists in the Pacific Northwest area, the HTDS principal investigator, members of the CDC scientific and media staff in Atlanta, and several members of citizen advisory groups in the Hanford region. We also examined various communication-planning materials that were made available by CDC.

Because the HTDS was an epidemiologic study with substantial public-health implications and because there was intense public interest in the Draft Final Report, the subcommittee felt that the study and the draft report should be thoroughly reviewed both for its technical aspects and for its effectiveness and balance in communicating to the public. We therefore went beyond the six questions that were posed by CDC and considered additional issues pertaining to scientific quality. Consequently, the main body of our report is not organized around the six questions from CDC, although responses to them are given in chapter 9 and summarized in the executive summary. The executive summary also provides conclusions and recommendations regarding the HTDS Draft Final Report.

The subcommittee's review of the HTDS Draft Final Report is organized around five main themes:

- Epidemiologic design and methods and clinical procedures.
- Dosimetry.
- Analysis of results.

- Statistical power and interpretation.
- Communication of results.

3

Evaluation of Epidemiologic and Clinical Methods

EVALUATION OF EPIDEMIOLOGIC METHODS

The HTDS investigators examined 3,441 persons who were born in 1940–1945 to parents living in the study areas near Hanford. Of these, 3,190 lived in the study areas at some time from 1945 to 1957, so thyroid dose estimates for them could be derived with the CIDER program developed by the HEDR investigators. Dose reconstruction used the available data and accepted methods.

The study design was appropriate to address the aims of the study. Among the possible designs, a cohort study with accompanying thyroid screening is optimal for minimizing the potential for bias. The investigators also chose the optimal populations to study: children who were young at the time of the greatest ^{131}I releases and who lived in the most highly exposed areas. Those two choices maximized the potential of the study to detect thyroid-disease effects. The interval between exposure and thyroid screening was adequate to allow radiation-induced thyroid diseases to become evident.

The epidemiologic methods were exceptionally good. The sample was based on an almost complete census of eligible subjects who were born in selected years and lived in what are believed to be the high-dose regions and subjects born in the same years who lived in regions where the doses from Hanford releases were lower. Efforts to locate subjects and to elicit study

participation were thorough. The investigators were able to locate 94% of the targeted sample—an excellent result, considering that 40–50 years had elapsed. Since the time of exposure they succeeded in communicating with about 98% of potential participants by telephone and had 15% refusals to participate. Thus, their losses due to nonlocation, noncontact and refusals totaled only about 23%. That is an excellent rate for a study that requires people to come in for examination, especially given their wide geographic dispersal. Furthermore, the losses to the study were comparable among those with high and low estimated doses, so participation rates were similar across the dose range.

There was a high level of quality control in the epiderniologic procedures. The interviewers were carefully trained by experienced interviewers, and detailed interview manuals were developed. Care was taken with data entry (double entry was used routinely), and range checks and consistency checks were implemented to help to reveal errors in the codes entered by the interviewers. Callbacks were used when there were clear errors or missing data.

Care was taken to maintain "blinding" in the clinical examination and other parts of the study where it could reasonably be done, to ensure that selection bias, interviewer-induced response bias, or clinical-examination bias would not creep in. For example, the residence-milk interviews with parents or the subjects to obtain residence and milk consumption-rate histories were conducted before, clinical examination so that neither subjects nor interviewers would know subjects thyroid status (except in the case of previously diagnosed thyroid disease).

In short, the HTDS was designed with great care to eliminate bias due to selection of subjects and due to reporting and detection of disease. This study compares favorably with other epiderniologic studies in these respects.

The HTDS considered an appropriate set of potential confounding variables, including sex; age at first exposure to ^{131}I from Hanford; age at examination; history of diagnostic,

therapeutic, and occupational radiation exposures; smoking history; ethnicity; and estimated thyroid dose from the NTS. A few other variables could have been considered, such as number of childbirths (for women) and family history of thyroid disease, but it seems improbable that these would have substantially altered the results because it is unlikely that they would be differentially distributed across the dose range. Another potential confounding variable was global fallout, which is discussed in chapter 4.

MILK-CONSUMPTION ESTIMATION

The consumption of fresh milk is an essential part of the pathway from release of ^{131}I to thyroid dose, so information on the amount and sources (for example, backyard cow versus commercial dairy) of milk consumed is important in estimating the individual doses. A questionnaire (called the CATI, for computer-assisted telephone interview, in the HTDS Draft Final Report) was developed to obtain information on residence history and milk sources and amounts during 1944–1957 from the mother or a surrogate (such as father or older sibling) who would presumably know about a subject's dietary patterns in childhood. The questionnaire was carefully developed and went through numerous revisions and multiple field tests. Professionals with expertise in questionnaire development commented on ways of eliciting questionnaire information to maximize the completeness and accuracy of recall. The development process was about as good as it could be.

Although the body of the HTDS Draft Final Report discusses how the milk-consumption questionnaire was developed and tested and attachments at the end of the report deal with it, there is not enough explanation of how open-ended information was coded. In addition, more needs to be stated about how much probing was used during the interviews, whether the booklet devised to stimulate memories of events that occurred 40 years earlier worked effectively and whether such memories were

checked for reliability in a subsample (for example, by comparing reports of the two parents).

The accuracy of Such information, obtained 40–50 years after the fact, is unknown. Food-consumption reporting that uses questionnaire-based methods is recognized as error-prone. Assessment (validation) of questionnaire efficacy in estimating true dietary intake generally requires careful study as a separate issue (Willett, 1990) before the questionnaire is used to relate food intake to risk of disease. The HTDS investigators were not able to evaluate the questionnaire's validity, nor did they cite any evidence from other studies about the validity of childhood milk-consumption reports decades after childhood. Milk consumption is often found to be among the better-reported elements of recent diet (for example, in the preceding year) when methods typical in epidemiology studies are used (Salvini and others, 1989), but we are aware of no studies that directly investigated the reliability or validity of retrospective milk-consumption reports by surrogates.

A few studies have considered self-recall of diet after considerable periods. Dwyer and others (1989), considering retrospective recall of childhood (age 5–9 years) food consumption by 72 middle-aged subjects who had originally been assessed as children, reported a correlation of 0.3 between the retrospective reports of whole-milk consumption and food histories taken during childhood. It seems credible that mothers' reporting of their children's early milk-consumption habits could be as good as the Dwyer group's findings or possibly somewhat better, but reliance on siblings or other relatives for information about early milk consumption seems unlikely to be better than the self-reports studied. Furthermore, a correlation of 0.3 does not indicate good predictiveness of childhood eating habits.

The HTDS investigators made a strong effort to obtain information from parents or surrogates. We would expect the mother to be the best source of childhood milk-drinking information. However, for 26% of the interviews, the investigators had to rely on information from some other family member.

Furthermore, because of parent deaths and other factors, they were unable to obtain any interview for 38% of the subjects. For the 38%, they obtained from the subjects themselves a residence history and at least the sources (backyard cow, commercial dairy, and so on) of milk drunk during 1944–1957, but not such details as the amount consumed. Thus, they had to use default values for a substantial fraction of the subjects, and this probably introduced measurement error into the data analyses.

The subcommittee's evaluation of the milk-consumption assessment is that the investigators did the best they could under the circumstances but that the resulting data have high intrinsic uncertainty. The effect of that uncertainty on the statistical power of the study is discussed in chapter 6 of this report.

The HTDS report has a substantial description of the collection of dose-related data from people but relatively little information on how these data were used. The input to the CIDER program is described as "scenarios", but these are not explicitly described, nor is their construction from the data. That is separate from how the CIDER program uses the scenarios to generate doses. There are a number of references to the use of default values in the CIDER program, but there is no discussion of which parameters used default values or of the degree to which default values changed as life circumstances changed for a given person (for example, if a person moved from a farm to a city).

EVALUATION OF CLINICAL-DATA COLLECTION

The clinical examinations and laboratory studies used the best modern methods of detecting and defining thyroid disease. Subjects were given physical examinations, including thyroid palpation by thyroid specialists, an ultrasonographic examination, and appropriate thyroid-hormone and thyroid-antibody blood tests. Quality control of the laboratory tests and ultrasonographic examinations was good. The clinicians were kept blinded to subjects' residential or milk-consumption history to avoid possible subtle clinical biases.

The study could have been strengthened if additional information on the following clinical and pathologic aspects had been gathered and presented:

- Added confidence is needed in the cytopathology data. The cytopathologic interpretation of thyroid fine-needle aspirations (FNAs) is a key step in distinguishing benign and malignant nodules. It would be useful to have the slides reviewed by two or three cytopathologists who are expert in thyroid disease and to have a consensus diagnosis when differences in interpretation are encountered. Special attention to the category of acellular and hypocellular aspirations that contain colloid is needed. They were categorized as benign, and that is not typical clinical practice. What constitutes an adequate biopsy should be defined in terms of numbers of cells and preparation technique.
- Some subjects had nodules that were not biopsied during the course of the study. For people with nodules greater than 1 cm in diameter at last observation, a followup, examination including an ultrasonography to look for nodule progression could be useful. If progression were detected, an FNA would be useful to document the characteristics of the lesion.
- A tabular presentation of the pathways to diagnosis would help readers to assess how the final diagnoses were assigned. For each clinical outcome, there is more than one way to make the assignment. For some, such as hyperparathyroidism, this is straightforward (either high calcium with high parathyroid hormone or a confirmed diagnosis before the study). For others, such as thyroid cancer, it is more complex (combinations of palpable nodule smaller than 1.5 cm, palpable nodule of at least 1.5 cm, nonpalpable nodule of at least 1.5 cm, FNA, surgery, prior diagnosis, cancer

in largest nodule, and cancer as an incidental surgery finding). A table for each diagnosis with a list of the methods of diagnosis and the number of each instance should be included in the fall report, and these data should be looked at for any indication of unsuspected ascertainment bias.

- The public has expressed concern that the HTDS analyses of the clinical data have "missed the forest for the trees"; that is, examining the fine categories of diagnoses might have caused the data analyst to miss trends, that occurred in broader categories of thyroid disease. Inasmuch as autoimmune thyroiditis, Graves disease, otherwise unexplained hypothyroidism, and ultrasonographic texture changes all are associated with autoimmune processes, one could score a person with any or all of these as positive for a new global variable of autoimmune thyroid disease (a broader category than the one by the same name in the HTDS). Similarly, the variable "any thyroid nodule", which was already analyzed in the HTDS report (tables VIII-53 to VIII-55), is a global variable of nodular thyroid disease. Any evidence of hyperparathyroidism or abnormal mineral metabolism could constitute another global category of disease. These variables would give a broad-brush view of thyroid disease in relation to ^{131}I dose, which would help to ensure that the fine diagnostic categories used in the HTDS report did not miss possible variations in broad categories of thyroid disease. However, a still broader category of "any thyroid disease" is not recommended, because combining pathophysiologically unrelated outcomes lacks biologic plausibility.

COMPLETENESS OF ASCERTAINMENT OF THYROID DISEASE

One possible weakness in the ascertainment of preexisting thyroid cancers results from the HTDS investigators inability to locate death certificates for all the decedents. Causes of death were not ascertained for 39 of 541 deaths. Given the small

number of thyroid cancers in the study (only 14 had estimated doses and could therefore be used in the dose-response analysis), a thyroid-cancer death among the unlocated cases could affect the results. This possibility cannot be ruled out, but it seems improbable that there would be any thyroid-cancer deaths among the 39 decedents for whom a death certificate or medical records where absent. First, there were no thyroid-cancer deaths among the 502 whose cause of death was known, so finding one among the 39 seems unlikely. Second, the probability of dying from thyroid cancer by the age of 55 is very small: the rate is about 4 per 100,000 persons among both males and females, and thyroid cancer accounts for every 1,000 deaths in the age range 0–54 years. Even if one triples the rate to account for a possible radiation effect, the probability is still very small that a thyroid-cancer death would be among the 39 whose death certificates were absent. Thyroid cancer could also be listed as a contributory condition on the death certificate, but this is rather unlikely unless it was part of the chain of disease leading to death, which again would have a low probability. A consultant to our subcommittee indicated that, of the 119 thyroid-cancer cases he had seen that were diagnosed before the age of 20 and had an average of 20 years of followup after diagnosis, only one led to death from thyroid cancer (Ernest Mazzaferri, personal communication); this further confirms the low likelihood of missed thyroid-cancer deaths.

Data on other thyroid diseases, but probably not thyroid cancer, can also be lacking because medical records are missing. The investigators reported that 37% of the 1,264 medical records they sought could not be obtained and that they were not able to obtain pathology or cytology slides for 10 of the 52 people on whom they sought them. It would be desirable for them to indicate for how many potential thyroid diagnoses they were unable to obtain any medical confirmation, preferably with a breakdown by reported type of thyroid disease.

SUSCEPTIBILITY FACTORS FOR THYROID DISEASE

At the various public meetings, several people who lived in downwind areas stated that they and their families had experienced more-frequent thyroid diseases than would have been expected in the population at large. They could be right, and their disease could have been the result of unusual fallout or ingestion patterns. However, thyroid disease does tend to run in families, and the particular occurrences could be related to genetic factors within the families, chance occurrences, or even mistaken diagnoses. The findings by several research groups that many of the thyroid cancers being found in Belarus and Ukraine, downwind of Chernobyl, have relatively unique *ptc3* mutations is one line of evidence for genetic factors in the disease following ^{131}I exposure. An enumeration and study of such clusters could have been undertaken but were not.

Thyroid cancer is not a common disease, and it would be reasonable in future epidemiology surveys to identify, document, and investigate clusters with molecular-biology probes to characterize genetic polymorphisms that could make people more sensitive to ionizing radiation or to look for oncogene prevalence in affected subgroups. These methods are developing rapidly, and will probably play a role in future environmental-epidemiology studies.

EVALUATION OF MORTALITY DATA

As part of the study protocol, death certificates for members of the Hanford cohort who had died were obtained. On the basis of those early deaths, the investigators calculated a standardized mortality ratio (SMR) for each cause of death. Overall mortality in the Hanford cohort was 20% higher than that in Washington state, which served as the reference population. The increase was due largely to deaths from congenital anomalies and "conditions in the perinatal period", although the SMR for cardiovascular disease was also somewhat increased. In an analysis

by geographic area, the highest SMR (close to 2.0) was observed in Franklin County.

Some of the initial findings raise suspicion about the contribution of radioactive releases from Hanford. The particular vulnerability of the fetus and infant to adverse effects of radiation is well known, and the mortality experience of this group in the Hanford cohort was somewhat unusual. Finding the highest mortality in the county that is closest to the Hanford site is also suggestive.

To follow up those observations, the investigators re-analyzed the data, breaking the total period into times before and after the peak releases of ^{131}I (between March and November 1945) and categorizing people accordingly. In the Draft Final Report, people are categorized by year of birth. A supplemental table later provided to the committee by the investigators categorized people by year of death as well. For congenital anomalies and conditions of the perinatal period, the year of birth and year of death are almost always the same. Table 3.1 compares the results of the analyses by year of birth and by year of death for the two causes of death.

Table 3.1 Standardized Mortality Ratios for Selected Causes of Death by Birth and Death Years

SMR (95% Confidence Interval)

	Year of Birth		Year of Death	
Cause of Death	1940–1944	1945–1946	1940–1944	1945+
Congenital anomalies	1.55 (1.06–2.19)	1.31 (0.63–2.41)	1.78 (1.15–2.63)	1.19 (0.69–1.90)
Conditions of the perinatal period	1.73 (1.33–2.21)	2.86 (2.09–3.83)	2.18 (1.68–2.78)	1.93 (1.41–2.59)

In both versions of the analysis—whether data are arrayed by year of birth or year of death—SMRs in the period both before the Hanford releases (before 1945) and after 1945 are elevated. For congential anomalies, the excess are slightly larger before 1945 than after, and the same holds true for deaths due to conditions in the perinatal period when considered by year of death. On the basis of the analysis, there appears to be no evidence that ^{131}I is responsible for the increased mortality from these causes in the Hanford cohort. However, when viewed by birth year, the SMRs for perinatal deaths do appear to be significantly higher in the period after the beginning of Hanford releases than before. The discrepancy between the two views of the same data is puzzling and needs further explication. There remain additional questions that the Hanford investigators could answer to resolve lingering questions:

- What is the distribution of types of anomalies and of "conditions of the perinatal period" in the two periods? The question of whether neural-tube defects are in excess around the Hanford site has been raised by another study (Sever and others, 1988) and the information available from the current study should be presented. The suggestion of a possibly larger excess in the 1945–1946 birth cohort than in the 1940–1944 birth cohort indicates that a more detailed presentation of the data is warranted.
- To what extent is the excess mortality associated with these two causes independent of the geographic excess in Franklin County?

AMERICAN INDIAN TRIBAL ISSUES

The subcommittee had some concerns about the extremely low number of American Indians in the study. The low compliance or cooperation might not be completely the fault of the

American Indian tribes. There are established methods for approaching and establishing rapport and trust with ethnic communities. The subcommittee wonders whether all these methods were tried. Perhaps a review of the methods used would be helpful in a future study.

A second concern is that American Indians had additional pathways of exposure to radiation from the Hanford site. For example, some tribes consumed much fish from the Columbia River. The river was contaminated by various radionuclides because it was used to cool the eight single-pass production reactors, in addition to the periodically increased ^{131}I that would have reached the river from rain and washoff from soil and vegetation. However, the thyroid dose resulting from aquatic pathways was likely to have been much smaller than that due to the consumption of fresh milk or leafy vegetables.

4

Evaluation of Dosimetric Methods and Results

This section presents a brief critique of HTDS dosinietnc methods and evaluates the quality of the dosimetric component. Our purpose is to assess whether there are any flaws in the estimation of the individual thyroid doses calculated by the HTDS, whether uncertainties are properly accounted for, and how much the NTS and global fallout might have affected the results of the HTDS.

BACKGROUND

One of the disputed aspects of the estimates of exposures of children downwind from the Hanford site pertains to the amount of ^{131}I released from the site, which in this case originated in the T and B reactors on the Columbia River; these reactors became operational in late 1944 or 1945. The fuel rods were "cooled" for various periods to let the short-lived fission products decay and then carried to separation plants, where the plutonium was ultimately reclaimed and the remaining ^{131}I effluent was a byproduct of the chemical separation procedures. The quantities of ^{131}I released to the air from the separation processes used to obtain plutonium from the uranium fuel rods of the reactors at the Hanford Atomic Product Operations were estimated, and meteorologic information and environmental-transfer models were used to assess the environmental concentrations of ^{131}I in air, milk, and other foodstuffs. Personal information on the whereabouts and

dietary habits of the study subjects led to the estimation of the thyroid doses from ^{131}I When the operation being described is 50 years in the past and the release and deposition of a radioisotope with only an 8-day half-life are being estimated, reconstructing the dose to the thyroids of a somewhat mobile population surrounding the effluent site is extremely difficult, even if it is based on the best records to be found and if due care is used in the calculations and interpretation of the scant data that are available.

In the HTDS Draft Final Report, the thyroid doses were calculated on the basis of information provided by the HEDR project. It is the subcommittee's understanding that the HEDR project estimated the daily environmental concentrations of ^{131}I (in indoor and outdoor ground-level air, in milk from a backyard cow and from commercial sources, in leafy vegetables, and so on) for the period from December 26, 1944, to December 31, 1957, and for geographic locations in the study domain where the subjects could have been exposed—from the Canadian border to central Oregon and from the eastern border of the Idaho panhandle to the center of the Cascade mountains. To do that, the HEDR project performed 100 realizations of its computer codes. In each realization, the input parameter values were randomly selected from subjective probability-distribution functions; 100 results (^{131}I, concentrations in air, milk, leafy vegetables, and so on) were provided for each day for each location of interest. Uncertainties were assigned according to histograms of the results.

Using the calculated ^{131}I environmental concentrations and the answers provided by the subjects as to their residence and dietary histories, the HTDS estimated their ^{131}I intakes via inhalation and via ingestion as a function of time. The products of the ^{131}I intakes and the ^{131} I thyroid dose coefficients for inhalation and for ingestion yielded the thyroid dose estimates. It is unclear to the subcommittee how the uncertainties in ^{131}I intakes were estimated by the HTDS.

One of the difficulties encountered in reviewing the work done by the HTDS is that the method used to calculate doses

is not clearly described in the documents that were provided to the subcommittee. This information is presumably scattered in the large number of documents that were prepared by the HEDR project and the HTDS in the course of their work. It is recommended that the HEDR project develop a single document in which the methods, assumptions, coefficients, and shape and magnitude of the coefficient uncertainties are summarized, the nature of the results provided to the HTDS is clearly described (with their strengths and weaknesses), and the ways in which the HTDS made use of the HEDR results to estimate individual thyroid doses are described in detail.

During the 1950s and early 1960s, two other environmental sources of ^{131}I contributed to the thyroid doses received by the populations around Hanford: the nuclear-weapons tests carried out at the NTS, mainly in 1952, 1953, 1955, and 1957; and the nuclear-weapons tests conducted outside the United States, which gave rise to "global" fallout, mainly in 1954, 1956, 1957, 1958, 1961, and 1962. Only the first of those two sources was taken into consideration by the HTDS. A crude assessment of the thyroid doses resulting from "global" fallout near Hanford has been prepared for the purposes of this review (appendix C).

HEDR CALCULATIONS OF ENVIRONMENTAL CONCENTRATIONS OF 131I FROM HANFORD AND RESULTING DOSES: ASSESSMENT OF KEY PARAMETER VALUES

Because of the paucity of ^{131}I environmental data, HEDR used a suite of environmental-transfer models to simulate the transfer of ^{131}I from the point of release to the dietary products consumed by the subjects and to ground-level air.

The HEDR models have been subjected to numerous reviews, for example, the Technical Steering Panel (TSP) and CDC review of the RATCHET model, the TSP external review of the environmental-accumulation dosimetry code development, two workshops held by CDC to address issues related to the dose and risk assessments, and a recent review prepared by this

subcommittee. In addition, the various computer codes have been tested by staff independent of the developers to ensure correct implementation of the models. The dose assessment has been found, on the whole, to be structurally sound for the estimation of thyroid doses to representative, hypothetical persons, but minor errors have been found and doubts have been raised about the validity of the results for particular environmental conditions and for the estimation of thyroid doses to specific people. In its letter report dated January 25, 1999, this subcommittee recommended that

> The HEDR investigators supplement their description of the model with an account of the origin of the errors made with regard to the estimation of the ^{131}I concentrations in pasture grass on the basis of measurements, the impact on the predicted values when the errors are corrected, and a preliminary assessment of the effect of reparameterization on estimates of absorbed dose to the thyroid. Until such a reassessment has been made it is difficult to know whether the current estimates of dose to representative individuals need recalculation.

As of September 1999, the subcommittee had not received any response to its recommendation.

Hourly releases of ^{131}I based on recorded data on the average reactor power levels and on the burnup of the fuel at discharge are provided in *PNWD-2033 HEDR* (volumes 1 and 2) (Heeb, 1993) for 1944–1947 and in *PNWD-2222 HEDR* (Heeb, 1994) for 1944–1949. For 1950–1972, only monthly release estimates are reported in *PNWD-2222 HEDR* (Heeb, 1994). The computer model for the source term (STRM) generated 100 realizations of the emissions for each hour for 1944–1949, which were aggregated into daily releases. The emissions were aggregated into monthly releases for 1950–1957.

The RATCHET model for environmental transport has a significant source of uncertainty built in that is not accounted for: the interpolation of the wind fields. This uncertainty is not spatially uniform. There should at least be a study of the sensitivity of the deposition estimates with respect to local errors in the wind field. According to Ramsdell and others (1994) (table 1), inverse-square-distance weighting was used for the interpolation. That method does not produce any information about the uncertainties associated with the interpolated values. The paper does not justify the choice of the method; in particular, it is known to be very sensitive to the location pattern of the data points. Inverse-square-distance weighting is known to be "isotropic"; that is, it incorporates the distance from the data points to the interpolation point but not the directions. In the case of wind fields, one would expect the direction from an interpolation point to a data point to be important. The algorithm is also insensitive to the directions between the pairs of data points; it incorporates only their relative distances from the interpolation point.

The environmental-accumulation model developed by the HEDR project (DESCARTES) provides the concentrations of ^{131}I in different plant and animal products for numerous space-time combinations, whereas the individual-dose model (CIDER) is used to estimate the ^{131}I body burdens and doses in humans of various ages and both sexes on the basis of ingestion and inhalation pathways. The equations used in the DESCARTES and CIDER models and the key input-parameter values and their distributions can be found in *PNWD-2023 HEDR* (Snyder and others, 1994). The DESCARTES code was operated with a daily time step for 1944–1949 and with a monthly time step for 1950–1957 (Farris and others, 1994; Price, 1994).

The thyroid dose via ingestion is due primarily to the consumption of milk produced by cows that fed on pasture grass.

For the purposes of this report, the thyroid dose from this pathway can be expressed as follows for an acute release of ^{131}I:

$D_m = (ST)(VEG/ST)(T_{veg})(PI)(F_m)(MDF)(MCR)(DCF)$,

where

D_m = thyroid dose, Gy,

ST = source term, Bq,

VEG = concentration of ^{131}I in pasture grass at time of fallout, Bq kg^{-1},

T_{veg} = mean time of residence of ^{131}I on pasture grass, d,

PI = pasture intake by cows, kg d^{-1},

F_m = transfer coefficient of ^{131}I to cow's milk, d L^{-1},

MDF = milk-distribution factor (which takes into account the fact that milk consumed might not be of local origin,

MCR = milk-consumption rate by subject, L d^{-1}, and

DCF = thyroid dose coefficient, Gy Bq^{-1}.

Those key parameters will be considered in turn. Most of the parameter values are presented in Snyder and others (1994).

Source Term:

The source term for 1944–1947, before the installation of removal devices for ^{131}I, appears to have been well estimated on the basis of the available historical data. It is during these early years that the largest ^{131}I releases occurred, but there is concern

that the ^{131}I releases for later dates have been underestimated by the HEDR project. The following are Hoffman's analysis of the HEDR estimates and not necessarily those of the committee. Several reasons for such an underestimation are listed in a draft report prepared by Hoffman and others (1999):

- The HEDR project estimates of the amounts of ^{131}I processed and released in 1959 and 1960 are substantially less than the amounts reported by Warren (1961), which are the source of HEDR release estimates.
- The HEDR project documents use Warren (1961) as a source but do not evaluate the credibility of his values; for example, he projected an unrealistically high scrubber efficiency.
- The HEDR project misapplied measured release-factor data: from 1959–1960 to the period 1951–1957, when less emission-control equipment was in place.
- The HEDR project incorrectly accounted for operation of the silver reactors in the B and T plants by inexperienced personnel during the first 18 months after installation in 1951.
- The HEDR project substantially underestimated the source-term uncertainties for the B, T, and REDOX plants.
- The HEDR project inadvertently used the medians instead of the arithmetic means of the monthly source terms for the air-concentration and ground-deposition calculations.
- The HEDR project did not propagate the source-term uncertainties to air concentrations, ground deposition, and doses.

The subcommittee did not investigate whether the reasons for a possible underestimation of the source term, as proposed in Hoffman and others (1999), are valid.

Hermann and Hermann (1996) evaluated the ^{131}I releases in 1948–1960; they based their estimates of emission-control efficiency on investigations conducted at fuel-reprocessing plant WAK in Karlsruhe, Germany, and concluded that the HEDR project had overestimated the efficiency of emission-control equipment for ^{131}I. They also concluded that the cooling time (time elapsed between fuel discharge from the reactor and fuel dissolution in the chemical plant) might not have been properly taken into account. Hermann and Hermann (1996) estimated an ^{131}I release of about 250,000 Ci in 1948–1960, whereas the HEDR estimate for the same period is about 54,000 Ci. In comparison, the ^{131}I release in 1944–1947, which seems to have been well estimated, was about 685,000 Ci. The estimates of the total releases of ^{131}I in 1944–1960 are therefore 740,000 Ci according to Heeb (1994) and 940,000 Ci according to Hermann and Hermann (1996).

Concentration of131I in Pasture Grass at Time of Fallout:

The concentration of ^{131}I in pasture grass at the time of fallout is characterized by the value of VEG/ST. It is derived from the daily ground-deposition densities of ^{131}I that are provided by the RATCHET code (Ramsdell and others, 1994). The deposition pattern is based on mathematical modeling, with little validation of the calculated deposition pattern of ^{131}I As shown by Napier and others (1994), there are discrepancies between the measured and calculated values of vegetation samples that have been collected since 1945. It might have been helpful to measure the ^{129}I pattern of deposition in the geographic domain considered by the HTDS. Because ^{129}I, a radioactive isotope of iodine with a very long half-life, behaves in the same way as ^{131}I and because Hanford was a dominant source of ^{129}I as long as the chemical plants were in operation, measurements of ^{129} I concentrations in soil would have

produced valuable information regarding the overall geographic deposition pattern of ^{131}I, even though the day-to-day temporal pattern of releases of ^{129}I, is not highly correlated with those of ^{131}I.

The derivation of the ^{131}I concentration in pasture grass from the ground-deposition densities makes use of a relationship that is valid only for dry deposition processes. It is not clear whether that relationship was used when precipitation occurred. Hoffman and others (1992) showed that the parameter values are substantially different for dry and wet processes. Because precipitation amounts in counties remote from Hanford are about 2–3 times higher than for counties near the site (see appendix C), this could have resulted in differences in a degree of overestimation or underestimation of the thyroid doses.

Mean Time of Residence of 131I on Pasture Grass:

The mean time of residence of ^{131}I on pasture grass seems to have been well estimated. Because of the short radioactive half-life of ^{131}I, the thyroid dose is relatively insensitive to the choice of value of this parameter.

Pasture Intake by Cows:

The values chosen for pasture intake by cows are not presented in Synder and others (1994), because they are extracted from a database described in Beck and others (1992) but later revised in an unpublished report submitted to the chair of the HEDR Technical Steering Panel (J.E. Till) in 1994. It would have been helpful to the subcommittee to have a report available on the topic.

Transfer Coefficient of 131I from Cow's Intake to Milk:

The central estimate of the transfer coefficient of ^{131}I from cow's intake to milk is taken to be 1.2×10^{-2} d L^{-1} for a dairy herd and 9.2×10^{-3} d L^{-1} for a single cow (Snyder and others, 1994). Those values seem to be too high by a factor of about 2,

compared with the central estimates used for NTS fallout—4×10^{-3} d L^{-1} in NCI (1997) and 4.2×10^{-3} d L^{-1} in Simon and others (1990) and with values obtained after the Chernobyl accident, which ranged from 2×10^{-3} to 6×10^{-3} d L^{-1}, according to Hoffman and others (1988). However, this is an open issue, in that the measured values of this transfer coefficient vary over a large range for reasons that remain largely unexplained.

Milk-Distribution Factor:

Data on the distribution of milk within the project domain are not presented in Synder and others (1994), because they are extracted from a database described in Deonigi and others (1994). It would be helpful to know how the uncertainties and the correlations in the distribution of milk were taken into account by the HTDS.

Milk-Consumption Rate:

The HEDR project selected milk-consumption rates for representative people in 12 age and sex categories (Anderson and others, 1993; Farris and others, 1994) on the basis of the results of the 1977–1978 Nationwide Food Consumption Survey and extrapolation back to 1945–1957. The HTDS used these rates when reliable information could not be determined from the personal interviews. However, how they were applied is unclear in the Draft Final Report. Because a number of assumptions are made (without quantifying the uncertainty) in Anderson and others (1993) and extrapolations are made from current food-consumption habits back to the relevant period, largely on the basis of broad-scale measures rather than measures particular to the people in the study, there will be at least an uncertainty associated with the extension of population or group practices to specific persons. It would be useful to know whether the milk-consumption rate distributions given in Anderson and others (1993) were used by the HTDS and whether milk-consumption rates were assumed to be constant in a

given age category or were taken to vary with age within an age category.

Thyroid Dose Coefficient:

The value of the thyroid dose coefficient depends on three parameters: the uptake of ^{131}I from blood by the thyroid, the time of retention of ^{131}I in the thyroid, and the mass of the thyroid gland. The values of those three parameters depend to some extent on the amount of stable iodine in the diet. The thyroid dose coefficient varies strongly as a function of age, the average value for infants being about 10 times greater than that for adults. That is caused by the variation of the mass of the thyroid gland as a function of age; the uptake of ^{131}I from blood by the thyroid and the time of retention of ^{131} I in the thyroid do not vary substantially with age.

The central estimates of the thyroid dose coefficients that are used by HEDR and the HTDS are taken from the International Commission on Radiological Protection (ICRP, 1990) and assume a sufficient amount of stable iodine in the diet; the uncertainties are taken from Dunning and Schwarz (1981). Those sources of information are generally accepted as the best available. However, it would be useful to know whether the thyroid dose coefficient was assumed by the HTDS to be constant within a given age category.

HEDR DOSIMETRY-MODEL VALIDATION

For an ideal validation of the dosimetry model, one would like to have measured doses for at least a subset of the people in the study; then one would compare the doses predicted for those people with the measured doses. But those data do not exist. Instead, the validation study used other people for whom some measurements were available. The validity of the comparisons hinges on the cross-applicability of various assumptions and parameter values between the test population and the study population. Shipler and others (1996), of the HEDR

project, acknowledge the lack of fully appropriate data for the validation exercise.

However, a substantial effort was made to validate the HEDR models with the data available. The model-validation exercise consisted of comparing computational-model estimates with limited historical field measurements and experimental measurements that were independent of those used to develop the HEDR models. The results are provided in *PNWD-2221 HEDR* (Napier and others, 1994). The following review focuses on the results of the model-validation exercise and on the selection of the values used for the key parameters.

The thyroid doses received via ingestion were essentially due to the consumption of cows' milk, resulting from the cows' intake of fresh pasture grass. The historical measurements that are the most interesting for testing the validity of the calculated ingestion doses are the thyroid burdens, the milk concentrations, and the pasture-grass concentrations of ^{131}I, in that order. Similarly, the historical measurements that are the most interesting for testing the validity of the calculated inhalation doses are the thyroid burdens and the ground-level air concentrations of ^{131}I.

Thyroid Burdens

Nearly 7,900 thyroid burdens of Hanford workers are available for the period June 1945–August 1946 (Napier and others, 1994). The report did not indicate how many workers were included in the 7,900 measurements nor how the sample of workers was drawn. It was assumed that all workers had environmental exposures and that occupational exposures did not affect the mean thyroid burdens. The mean monthly thyroid burdens were lower by a factor of 2–4 than the values calculated by the HEDR model for 1945 but were in agreement with the calculated values for 1946. The calculated values were obtained by using the assumptions that the workers lived in Richland and that the milk that they consumed originated in the area around

Sunnyside. Under those conditions, the HEDR investigators believed that most of the thyroid burden was due to inhalation.

In view of the importance of the thyroid-burden data set to validate the HEDR models, it seems that a larger effort could have been devoted to the comparison of the observed and calculated values. For example, whether all workers actually lived in Richland and drank commercial milk originating in Sunnyside could have been checked. If it was true, the variability of the observed values, which is not provided in *PNWD-2221 HEDR* (Napier and others, 1994), could have been used to estimate the random uncertainties in individual doses, and research could have been carried out to investigate the bias observed for 1945. If not, the data set could have been stratified according to place of residence or origin of milk, and calculations done for each subgroup. For people with many thyroid-burden measurements, it also would have been worth while to analyze the temporal variation of the results.

Two more thyroid burdens are available for 1963. An acute, inadvertent release at the PUREX plant in September 1963 was extensively studied. In particular, pasture grass and milk at two farms—called Farm A and Farm B—were measured regularly, and thyroid counts were taken on two children who were consuming milk from a family cow at Farm B. The thyroid doses that were calculated for the two children with the HEDR models are in very good agreement with those derived from the thyroid burdens: there is a difference of only about 20% between the "measured" doses and the medians of the calculated doses. However, because the thyroid counts were taken 1 1/2 months after the PUREX acute release, it is likely that a fraction of the thyroid burdens arose from releases that occurred between the PUREX acute release and the time when the thyroid counts were made (Goble, 1968).

Milk Concentrations

The only milk concentrations reported in *PNWD-2221 HEDR* (Napier and others, 1994) are those measured at Farm A and Farm B in September 1963. The measured milk concentrations at the two farms differed by a factor of about 3. The calculated milk concentrations were the same at the two farms, and they were included in the same calculation cell, although Farm A was on the western edge of the cell and Farm B was on the eastern edge. The measured and calculated milk concentrations were in very good agreement for Farm A, whereas the measured milk concentrations in Farm B were about 3 times higher than the calculated milk concentrations.

As indicated in appendix C, concentrations of ^{131}I in milk were measured around Hanford as early as 1958. It would have been valuable to investigate how the calculated concentrations agree with the measured values.

Concentrations in Vegetation

Historical measurements of ^{131}I in vegetation at the Hanford site were available to the HEDR project for the period beginning in the middle of 1945 and continuing to the present. For example, over 3,500 samples were reported for 1946. The samples in which ^{131}I was measured were most often labeled "vegetation", but it was noted that the samples were usually sagebrush.

Sagebrush is very different from pasture grass, so these measurements might not be very useful for deriving the concentrations in pasture grass. And errors were made because wet weights were sometimes used instead of dry weights; these errors are being corrected by Pacific Northwest National Laboratory, which will issue a revised analysis. The subcommittee will review the revised analysis when it is available.

Concentrations in Air

Relatively recent Hanford data on atmospheric dispersion include a dataset of coupled monthly source terms and

environmental measurements of atmospheric krypton-85 (^{85}Kr). The number of locations sampled per year increased from four in 1984 to 11 in 1987; up to 12 samples per year were taken from each location. Detailed results were presented for three of the sampling stations that were in the same grid cell (North of Hanford Site 300 Area). Air concentrations were calculated from two sets of meteorologic data: those available in the 1940s and those, more complete, available in the 1980s. The calculations were in good agreement with the measurements when the 1980 meteorologic data were used and were about 3 times higher than the measurements when only the meteorologic data available in the 1940s were used.

Much less detailed information is presented for the other sampling stations where ^{85}Kr was measured. It would have been valuable to present the results in the same fashion as in the "North of Hanford Site 300 Area" grid cell to identify possible location-related biases in the calculated values.

Conclusions

On the whole, the results presented are good, inasmuch as there is agreement between measured and median calculated values ^{85}Kr air concentrations could have been analyzed more thoroughly. In addition, errors in the assessment of the concentrations in vegetation should be corrected, and many isolated data were not used in validating the models.

Possible Underestimation or Overestimation of Doses

Hoffman and others (1999) and Hermann and Hermann (1996) have reviewed the HEDR dosimetry and claimed that thyroid doses were substantially underestimated by the HEDR project. However, the I divergences that they identified occurred during years when ^{131}I releases were rather small, and their findings and HEDR results differed by only about a factor of 1.3 in total releases.

If doses were underestimated, it is important to consider the implications. To the degree that doses are underestimated, the imputed risk estimates will be too large, but systematic, across-the-board dose underestimation does not alter the statistical significance of dose-response trends. Hence, in the HTDS—in which, as it turned out, the primary issue became whether there is any association between ^{131}I exposure and thyroid diseases—the effect of possible dose underestimation might be minimal. If there was systematic underestimation of doses, the true statistical power of the study would be greater than one would estimate it to be given the dose distribution used by the HTDS, in which case negative results would be more persuasive than they are. However, if the doses were underestimated more for some study subjects than for others or if the effect of dose underestimation was greater for some subjects than for others (for example, because some are younger than others), variation in the underestimation or in its effect would act as another source of measurement error and tend to cancel out the gain in statistical power achieved by having generally higher doses. Therefore, a simple generalization about the effects of dose underestimation cannot be given.

In contrast, the HEDR project might have overestimated the transfer coefficient from cow's intake to milk by a factor of 2, and some of the validation comparisons also suggest an overestimation of doses by the HEDR calculations. The effect of dose overestimation could have implications for the interpretation of HTDS results because overestimation would mean that the statistical power of the study was also overestimated; this would imply that the negative results were not as definitive as the investigators asserted.

Doses might have been underestimated in some areas and overestimated in others. Possible reasons are errors in the deposition pattern of ^{131}I due to inaccuracies in wind flows or precipitation amounts and errors in the milk-distribution pattern.

It is beyond the mandate of the subcommittee to assess the direction and magnitude of possible dose misestimation

definitively. As stated in a previous letter report (National Research Council Letter Report to CDC: HEDR—Issues Regarding the Hanford Environmental Dose Reconstruction (HEDR) Atmospheric I-131 Pathway Models dated January 25, 1999), the subcommittee found the dose-reconstruction effort to be, by and large, a credible effort, although there were unresolved questions and considerable uncertainty in doses because of the lack of adequate historical measurements and other relevant data.

DOSE UNCERTAINTIES

Considerable efforts were made by the HEDR project to model the uncertainties attached to estimated thyroid doses. Uncertainties for many parameters used in the environmental pathways and radiologic dose modules (DESCARTES, CIDER, and CRD codes) are listed in Snyder and others (1994). The HEDR investigators assigned probability distributions to the various parameters on the basis of their assessment of the uncertainties in the parameter values. They then obtained 100 estimates based on the computer codes. In each estimate, for each parameter the values were randomly selected from the probability-distribution functions, so 100 results (^{131}I concentrations in air, milk, leafy vegetables, and so on) were generated for each day for each square in their grid that defined a location in their geographic domain. Uncertainties were assigned according to the frequency distributions of the 100 results.

A key point here is the possible distinction between variability (added to the data) and uncertainty inherent in the data. If one has an estimate for a parameter value and adds a random component, such as meteorologic data (Ramsdell and others, 1994, p. 571), this does not characterize the uncertainty in the original estimate (although the authors claim that it does). Adding the random component does not result in a random component with the same distribution as the added part, but rather is the convolution of the distribution of the original and the distribution of the random component. Regrettably, the distribution of the

original estimate, the distribution of the error of estimation, is unknown.

In other places in the publications mentioned above, the authors indicate that they have sampled from a subjective distribution, as opposed to using the estimate plus a random component. Aside from the question of whether a proper subjective distribution has been chosen, this approach is more appropriate. However, it is difficult to quantify the uncertainty associated with choosing or fitting the subjective probability distribution, and the authors have not incorporated this uncertainty into the calculations.

There might be adequate explanations for these discrepancies that do not appear in the publications or in the Draft Final Report. The omission and the lack of clarity of what the HEDR project did constitute inadequacies in communication and perhaps in scientific methods. Without adequate explanations in the Draft Final Report, the committee cannot confirm that the propagation of error in the dose reconstruction (the source term in particular) was the best choice of methods and is correct. This situation adds to our overall concern about the integrity of the uncertainties of the dose estimates and how they could affect the power calculations. By "integrity of the uncertainties" the subcommittee means the inclusion of all of the uncertainties and then understanding and using them in an appropriate way in the epidemiologic analyses.

It is not clear to the committee what was done by the HEDR project about the parameters that are not listed in Snyder and others (1994) or about the model uncertainties. For example,

- According to Hoffman and others (1999), the HEDR project did not propagate the source-term uncertainties to air concentrations, ground deposition, and doses for the 1950–1972 period. Evidence of uncertainty propagation could not be found in the HEDR reports made available to the subcommittee.

- No information could be found on the uncertainties related to the cows' feeding practices or to the commercial distribution of milk.
- No information could be found on the correlations between parameter values.
- No information could be found on the uncertainties related to the use of models to calculate thyroid doses from exposures.

The uncertainty estimates attached to the HTDS thyroid doses appear to be too low. The geometric standard deviations (GSDs) of some of those doses are estimated to be about 1.4. The GSD of the thyroid dose-conversion factor alone is taken to be 2.0 (Snyder and others, 1994), so, the GSDs of thyroid dose estimates should be at least 2. The authors of the report should explain why GSD values of less than 2 are found for many HTDS thyroid-dose estimates. It is to be noted that the GSDs of the thyroid doses calculated by the HEDR project are 2 or greater (Farris and others, 1996). An increase in the GSD from 1.4 to 2.0 quadruples the width of the 95% confidence interval.

The HTDS did not take into account the uncertainties associated with the recall—5 decades after the period of exposure—of the origin of milk, the milk-consumption rate, or changes in residence. That is a serious flaw of the methodology, which is discussed further in chapter 6.

In summary, although the subcommittee believes that the methods used by the HEDR project to estimate thyroid doses and their uncertainties are on the whole structurally sound, there remain quite a few errors or omissions that should be corrected, and supplementary work is desirable. In a National Research Council letter report dated July 27, 1998, the committee stated that the resulting dose estimates are highly uncertain, and the use of self-reported information in the construction of dose estimates

presents an opportunity for bias, such as overreporting or underreporting of activities that might substantially affect exposure estimates. The investigators are urged to evaluate further the use of some of the interview data to corroborate the HEDR exposure estimates and the default options. Modeling has inherent limitations that need to be acknowledged candidly. Given the insufficient data available on Hanford, there will always be considerable imprecision in the individual doses, and it is unlikely that further refinements in the model will increase the precision of these doses substantially. It warrants noting, too, that although the uncertainties in some of the components of the model have been accounted for, this is not true for all of them, and the total uncertainty is likely to be greater than has been publicly stated.

This subcommittee stresses again that a review of the dosimetry work conducted by the HEDR project and the HTDS is difficult in the absence of a single dosimetry document in which the HEDR work is summarized, the nature of the results provided to the HTDS is clearly described (with their strengths and weaknesses), and how the HTDS made use of the HEDR results to estimate individual thyroid doses is described in detail. In the absence of such a document, however, it seems that the uncertainties in the thyroid doses were underestimated by the HEDR project and consequently by the HTDS.

^{131}I FROM THE NEVADA TEST SITE AND GLOBAL FALLOUT

During the 1950s and early 1960s, two other environmental sources of ^{131}I contributed to the thyroid doses received by the populations around Hanford: the nuclear-weapons tests carried out at the NTS, mainly in 1952, 1953, 1955, and 1957; and the nuclear-weapons tests conducted outside the lower 48 states of the United States, which gave rise to so-called global fallout, mainly in 1954, 1956, 1957, 1958, 1961, and 1962. If global or NTS fallout resulted in significant thyroid doses in the counties used in the HTDS, and if the variation in doses among counties used in the HTDS was large for the sources of additional

exposure, then it would be important to take the global and NTS fallout into account in the HTDS. Only the first source was taken into consideration by the HTDS.

For the purposes of this review, a crude assessment of the thyroid doses resulting from global fallout near Hanford has been prepared in order to obtain preliminary evidence as to whether global fallout is likely to be an important confounding variable in the analyses pertaining to Hanford fallout and thyroid disease. Details of the assessment are provided in appendix C. The following paragraphs present a brief summary of the methods used and the results.

The global tests were carried out primarily in the Pacific at Bikini and Eniwetok and in the Soviet Union at Novaya Zemlya. The total fission yield from the tests was over 150 megatons, compared with a total of about 1 megaton for the tests carried out at the NTS. However, fission yield is only a crude indicator of the fallout deposition in the continental US because about 80% of the debris was injected into the stratosphere, where ^{131}I decayed and therefore did not contribute to human exposure. Of the 20% of the debris injected into the troposphere, a considerable fraction was probably deposited locally or regionally; the remainder was deposited primarily in the latitudinal band of the location of the test site and was mainly associated with rain. It would be extremely difficult with atmospheric-dispersion and-deposition models to predict with reasonable accuracy the fallout deposition at a given site in the lower 48 states on the basis of the yield of a particular test. The dose-assessment methods used were based on the sparse fallout data and on the small number of ^{131}I concentrations in milk in the Hanford area for some of the testing period that are available.

The basic methodology used by Beck (see appendix C) to estimate the ^{131}I doses near Hanford was as follows:

- For 1961 and 1962, the doses for Franklin, Adams, and Benton counties were estimated directly from the measured ^{131}I

in milk at farms near Hanford after the milk data were corrected for Hanford plant contributions. All three counties are part of the same milk shed and receive similar amounts of rain, so the milk-concentration estimates were assumed to apply to all three counties, and a single set of dose estimates was made for these three counties. (The doses in Yakima and Kittitas Counties, which are in the milk-producing area, would also be similar to those in Franklin, Adams, and Benton counties because of the similar rainfall pattern.)

- The strontium-89 (^{89}Sr) depositions in 1957, 1958, 1961, and 1962 were then estimated from the ^{89}Sr depositions at sites in the western United States and the measured monthly rainfall in counties in the Hanford area. The ratios of the deposition in Walla Walla and Stevens counties to deposition in Franklin, Adams, and Benton counties were used to estimate the milk concentrations in those counties in 1961 and 1962 from those measured in milk near Hanford.
- The calculated deposition in 1958 relative to 1961 and 1962 was then used to estimate the relative concentrations in milk in 1957 and 1958 in three counties.
- The ^{89}Sr deposition in 1956 and 1954 relative to 1958 was estimated from the gummed-film data, and the concentrations in milk were assumed to vary with the estimated deposition.
- Finally, the thyroid doses in each county were estimated by using the conversion factors for milk concentration to dose that were used for that county in the National Cancer Institute study (NCI, 1997) for the estimation of ^{131}I thyroid doses resulting from tests carried out at the NTS.

The thyroid doses from global fallout calculated as described above for the Hanford area are lower but of the same order of magnitude as those from NTS fallout, as shown in table 4.1. Considering the uncertainties in both the NTS and global fallout dose estimates, the differences between the two sets of dose estimates are probably not statistically significant. Thyroid doses for infants, children, and teenagers are also provided, by year of fallout and by county, in appendix C.

Table 4.1 Comparison of average estimated 131I doses (mGy) from Hanford, the Nevada Test Site (NTS), and global sources for study counties

County	HTDS	Global fallout[a]	NTS fallout[b]
Benton	172	3.8	12
Franklin	248	3.8	11
Adams	169	3.8	12
Walla Walla	86	8.7	30
Ferry-Stevens	39	11.8	14–19
Okanogan	11	not available	11

[a] Estimated doses for teenagers (see appendix C).
[b] Estimated doses for children born January 1, 1945.

In theory, global fallout is a potential confounding variable in the analyses of thyroid-disease risk posed by Hanford fallout. But in reality, the degree to which global fallout could confound the analytic results appears to be limited. The range of variation in global fallout between the Hanford high-dose and low-dose counties is small compared with the variation in Hanford dose (see table 4.1). In addition, the global-fallout exposures occurred during the teenage years and early 20s for the study population. Data from the best studies of external radiation exposure to the thyroid strongly indicate that the thyroid is much more sensitive to radiation-induced cancer in childhood, especially before the age of 5 years, than in adolescence or adulthood (Ron and others, 1995).

Hence, the effect of the Hanford exposures on thyroid-disease risk would be expected to be much greater per unit dose than exposures from the NTS or global fallout. Nevertheless, since no actual analysis of the impact of global fallout on the results has been done, the committee recommends that the HTDS investigators do so.

OTHER SOURCES OF RADIATION EXPOSURE

Ideally, estimates of the diagnostic and therapeutic medical radiation exposures received by each person in the HTDS should be factored into a dose-response assessment. The investigators did include questions about history of substantial medical radiation exposures in the questionnaire administered to subjects. In a supplementary set of analyses, they evaluated whether those exposures were confounders or effect modifiers of the dose-response results for ^{131}I versus thyroid diseases. However, they did not state how the radiation exposures were coded for analysis or how reliable the responses were, and this could be problematic: a crude, dichotomous coding or unreliable reports might adjust only partially for potential confounding. Nevertheless, it is not very likely that medical radiation exposure would be a confounding variable, in that the frequency and intensity of such exposures would have to be correlated with the magnitude of the Hanford ^{131}I doses for confounding to occur.

The thyroid doses from the medical radiation exposures asked about would probably vary appreciably among medical procedures and radiological practices. For example, in the 1940s and 1950s, the radiation exposure from taking a full-mouth diagnostic x-ray picture was often several hundred milligrays to the cheek. The machines often had poor collimation (Budowsky and others, 1956; Nolan, 1953) (that is, there was appreciable scatter of the radiation), and the adequacy of neck shielding was probably variable. Hence, some persons could have received thyroid doses from diagnostic dental procedures that exceeded those from Hanford fallout.

Questionnaire information on historical exposures is usually insufficient to impute thyroid doses to individuals, so one would prefer to have documented medical radiation doses to use in the analyses. However, it would be impossible to retrieve records to document many radiation exposures that occurred 2–5 decades ago, so it would not be feasible to estimate lifetime medical radiation exposures. In lieu of that, the investigators did list (in table VIII-46 of the Draft Final Report) the reported medical diagnostic and therapeutic radiation exposures for each thyroid-cancer case, and this is informative.

5

Evaluation of Statistical Data Analysis

MODE OF PRESENTATION OF RESULTS

The basic objective of the statistical analysis was to determine whether thyroid disease has increased among persons exposed to ^{131}I released from Hanford in the period under consideration. That objective was appropriately addressed by modeling the relationship between dose and the probability of occurrence of a thyroid disease. In particular, the relationship was modeled as a linear function—that is, a regression equation—by using the median of the 100 dose estimates for each person as his or her assumed dose. The HTDS investigators used these linear models in dose to model probabilities of disease and the numeric values of blood concentrations of various biomarkers. The use of linear models (or linear-quadratic models) for probabilities is common in radiation epidemiology and is generally based on biologic or radiation-protection considerations. Such models, however, can present difficulties in estimation, in that negative probabilities are not allowed to occur but can appear during the iterative procedure used to produce maximum likelihood estimates, especially if a generally negative dose-response relationship is evident in the data. In such cases, the models are said to have had problems in "converging". The HTDS investigators also used another approach: logistic regression. Logistic regression is a nonlinear model in which probabilities are never allowed to reach zero, so convergence problems are less common during maximum

likelihood fitting. However, the parameter estimates from logistic regression are arguably less easily interpretable in the radiation-biology or radiation-epidemiology setting because log odds ratios, not disease probabilities, are being modeled as linear in dose. The HTDS investigators used the linear models as their primary method of analysis but also gave the logistic-regression results, especially when the linear models failed to converge. In their analyses, convergence apparently was always achieved with the logistic models, but many of the linear models had convergence problems. The HTDS investigators clearly considered the linear models as the primary method of analysis; for example, power calculations were given only for linear models (section V). We focus most of our comments on their use of the linear models, but much of our critique is also applicable to logistic regression.

A zero slope in the regression equation indicates no association between dose and the probability of occurrence of a thyroid disease. Standard statistical tests were used to determine whether the slope was significantly positive. Because the investigators assumed that the association, if any, would be in a positive direction, they appropriately used a one-sided statistical test. For most of the thyroid diseases considered in the HTDS, the conclusion was that the null hypothesis of zero slope could not be rejected, that is, there was no clear evidence of a thyroid-disease effect due to exposure.

In the HTDS Draft Final Report, there was an overreliance on the maximum likelihood fitting of the linear dose-response model. For several of the important outcome variables, such as thyroid carcinoma, the model calculations failed to converge. A better organization of the results would have been achieved by expanding the tables of high-versus low-dose results (section VIII, pages 107–124) into quartiles or quintiles, so that disease and abnormality rates would be given for four or five categories of dose, and incorporating them into the presentation of the earlier results. In cases where the results of the linear model calculations as described failed to converge, it would then be

possible to rerun the analysis, using the average value of dose in each category as the predictor variable. That would probably have resulted in successful convergence and would retain reasonable power to detect an effect. Another approach would be to replace the maximum likelihood fitting with an ordinary least-squares analysis when the model failed to converge with maximum likelihood. Unless a large proportion of the estimated probabilities lie outside the unit interval (0,1), the slope test statistic is reasonable for testing for the presence of a dose-response relationship. For models in which convergence was still not obtained, it would be reasonable to report the value of the slope parameter at the point where the constraint that all outcome probabilities be positive was first violated. A confidence limit based on the profile likelihood for the slope parameter could have been calculated and would have been helpful, especially for comparison with other studies.

Rather than reliance exclusively on analyses that used the putative individual doses, an additional set of confirmatory analyses would be valuable. The basic parameters that define a person's dose are geographic location, source of milk (backyard cow, commercial milk, and so on), and amount of milk consumed. Analyses of thyroid-disease rates according to those basic dose-related variables would provide assurance that doses were not seriously misestimated and could further confirm (or contradict) the principal negative results.

Some results were presented in an abstract, rather uninformative manner. For example, there was a scatter plot of individual thyroid doses on a logarithmic scale but no table of frequency distribution of doses, which would have been much more useful. Similarly, one expects radiation-epidemiology reports to include tables that show observed and expected numbers of disease outcomes according to dose groups; these key tables were absent from the Draft Final Report.

Too little descriptive material supporting the results of the analysis was presented. A description of the estimated dose

distribution (distribution of median doses for people in the study) and disease frequencies and prevalence according to such important categories as sex, geostratum, year of birth, and amount of milk consumption in childhood would be helpful, especially in interpreting the finding that people in the least exposed geostratum appeared to have the highest rates of many of the thyroid diseases or abnormalities.

TYPES OF ANALYSES AND RELATED DOSIMETRY-ERROR ISSUES

The model of the dose-response relationship was given in the HTDS Final Draft Report (section VII, page 10) as

$P_j(d) = A_j + Bd$,

where

j = sex,

d = cumulative dose to thyroid,

$P_j(d)$ = probability that person of sex j and dose d has disease or condition in question,

A_j = baseline risk (risk without radiation, which can depend on sex), and

B = regression coefficient on dose (the slope of dose-response regression line).

There is sizable uncertainty in the doses reconstructed for individuals based on residential and especially dietary histories, and variations related to source term, meteorologic uncertainties, pasture deposition, milk concentrations of ^{131}I, source of milk, and

iodine metabolism need to be taken into account. It seems clear that analyses need to address those uncertainties explicitly and that the confidence intervals and the strength of the conclusions have to reflect them. That implies assumptions pertaining to the distributions of two error terms, E_1 and E_2:

$P_j(d) = A_j + B\,(d + E_1) + E_2$,

where

E_1 = error in estimating doses, and

E_2 = error in response to given dose.

The statistical-methods section (section VII) of the Draft Final Report described a model incorporating uncertainties in dosimetry, but the analyses (section VIII) used a simpler model that did not include dose uncertainties. Furthermore, no assumptions were stated for E_2, and little attention was given to the uncertainty represented by it.

As described in chapter 6, below, it appears that dosimetry-error issues were not fully treated in the analysis of the power of the HTDS. Ignoring dosimetry errors can lead to unrealistically narrow estimates of the confidence limits that are applied to the estimated parameter values. The statistical-methods section does describe a method for computing likelihood-ratio (LR) statistics when (as is true for the HEDR doses) errors in the dose estimates for each individual are correlated (section VII.C.3); however, this method is not used in the results section. The suitability of the LR method that the investigators presented to account for dosimetry errors depends on the validity of the Berkson model for errors (see chapter 6) and on accepting that the correlations between dose estimates are fully specified by the HEDR simulations. Despite those cautions, results from the LR approach could be useful. Although it is very unlikely that the

estimated dose-response relationships would change in an important way, confidence intervals that take dosimetry errors into account would provide more appropriate information about the uncertainty of estimates. It is recommended that confidence intervals be calculated.

ANALYSIS OF POTENTIAL CONFOUNDING OR EFFECT-MODIFYING VARIABLES

It was not clear from the Draft Final Report how confounders of dose-response relationships were treated, and results adjusted for possible confounders were rarely given. The HTDS investigators conducted analyses of the various thyroid-disease end points to evaluate a number of possible risk factors for confounding effects or effect modification (section VIII.D.20) but presented no tables to show a summary of the results of their analyses for any of the end points. Several such tables should be added to the report.

ASSUMPTION OF EQUIVALENT RADIATION EFFECT FOR MALES AND FEMALES

An important assumption in the main analyses was that rates of thyroid disease might differ between sexes in the absence of ^{131}I but that the radiation effect (which was calculated as an excess absolute risk) would be comparable for males and females. A number of studies have found that excess absolute risks posed by external radiation are greater for females than for males with respect to thyroid cancer (Ron and others, 1995) or thyroid nodules (Nagataki and others, 1994; Ron and others, 1989; Wong and others, 1996), so the assumption of comparability between sexes is a key assumption to be tested. The investigators stated that they tested it but presented no results for the reader to examine with respect to this assumption for any of the disease outcomes. An analysis that allows for differences between sexes in dose-response slopes should be presented.

ANALYSES BY AGE AT EXPOSURE

The prevalence of thyroid cancer induced by radiation depends heavily on age at exposure, so it would have been helpful to see a table showing dose-response analyses for those who were younger versus older in 1945, the time of the greatest ^{131}I irradiation, to examine whether there were indications of a radiation effect among those exposed at the lowest ages. In particular, it is recommended that results be presented for those exposed in utero and during the first 2 years of life. Likewise, because the magnitude of thyroid doses from ^{131}I fallout from the NTS and from global fallout was not greatly different from the Hanford doses in many study subjects, tables showing the results of analyses stratified by magnitude of NTS or global fallout are potentially important.

OUT-OF-AREA ANALYSES

The HTDS investigators took care to examine the results for the out-of-area participants, those who proved never to have been in the dosimetry area during the time of ^{131}I exposure. They performed sensitivity analyses in which the out-of-area participants with disease were assigned either the minimal (zero) or maximal (at the dose-assessment area boundary) likely dose and those without disease were assigned the converse. The two contrasting analyses test the minimal and maximal contributions, respectively, that the out-of-area subjects could make to the dose-response analyses. Either way, the overall results were essentially unaffected, and this indicates that their deletion from the main analyses did not produce a substantial bias. The effect of these cases was small probably because only about 7% of the subjects were out-of-area and the assigned doses for these subjects in the sensitivity analyses were relatively small (8–51 mGy).

However, the HTDS investigators made no attempt to model the out-of-area doses for persons who were included in the main analyses. That is, if persons were in their dose-assessment

area for only part of the time when there were ^{131}I releases, their doses were calculated only for the time when they were in the area. The investigators implicitly assumed that the dose was zero for any time when a person did not reside in the area. That assumption might or might not have been valid for some individuals, but no attempt was made to improve on the approach or to conduct a sensitivity analysis to evaluate how the assumption could have affected the results. That approach could have led to attenuated or biased results in that it estimated the total Hanford fallout doses for some people and only partial doses for others.

There was not even a tabulation of the fractions of the dose-modeled persons that were partly in and partly out of the dose-assessment area during the exposure period or, what would have been better, what fractions of them were out-of-area during the period of heaviest exposures (1944–1947), out-of-area only during other exposure periods, and entirely in-area. The committee cannot evaluate the potential for attenuation or bias by this factor without at least some information on its frequency, and we recommend that the issue of partial out-of-area HTDS subjects be examined.

GEOSTRATUM VERSUS DISEASE

The HTDS investigators examined thyroid morbidity according to geographic areas, which they called "geostrata". Given that outcomes (disease or abnormalities) appeared to differ by geostratum, an alternative analysis that stratified by geostratum. would be natural to consider. It would be difficult for thyroid carcinoma (owing to the few cases detected), but many of the other outcomes could be analyzed so that the dose-response relationships were estimated for the individual geostrata and then combined to yield a pooled dose-response estimate. Additional analyses are presented that are based on excluding the Okanogan and Ferry-Stevens geostrata; this could well have effects on the dose-response estimates similar to those of a stratified analysis, but one cannot be sure from the writeup. A set of analyses stratifying on

geostrata seems needed because the tabulations show that the disease-rates tended to be higher in areas with low fallout; this means that the geostratum differences would induce a negative association between ^{131}I and thyroid-nodule rates. It is recognized that it can be tricky to conduct an analysis controlling for a variable that is correlated with dose, because one does not want to control (remove) a large fraction of the variability in dose; but in this case, when it appears that geostratum is a potent confounder of the dose-response association, it seems necessary. Perhaps a judicious collapsing of similar geostrata can minimize the potential for "overadjustment" (Day and others, 1980) of the exposure variable.

Faced with a similar problem in the study of Utah NTS fallout and thyroid disease, Kerber and others (1993) conducted their primary analysis with stratification on coarse geostrata (by state), examining the association of thyroid neoplasms and ^{131}I dose within geostrata. It is recommended that the Hanford investigators perform a similar type of analysis to examine the possible association of thyroid nodules and other thyroid diseases with ^{131}I dose. This would provide assurance that a possible confounding variable had been sufficiently evaluated, either to ensure that a positive association was not masked by the geostratum variations or to detect a masked association.

GENERAL-POPULATION COMPARISON AND SCREENING ISSUES

When one takes into account the different contributions of ^{131}I from Hanford, NTS, and global fallout from weapons testing, everyone was exposed, so it was not possible to identify an unexposed control group. Concern has been expressed that a study in which everyone is exposed is not valid—that an unexposed group is needed to assess the risk posed by Hanford ^{131}I fallout. However, under the weak assumption of a monotonic dose-response relationship (that is, other things being equal, the larger the dose the greater the thyroid-cancer risk), it is not necessary to have an unexposed control group to estimate the risk. The slope of

the dose-response curve would provide a valid index of the risk even without an unexposed control group, provided that there is a sufficient range of doses and that the doses are estimated with reasonable accuracy. Problems in trying to define and use an unexposed control group are discussed below.

The primary analyses of the cumulative incidence of thyroid cancer or other thyroid conditions were dose-response analyses of the subjects in the study. These analyses are appropriate to address the scientific questions regarding the association between ^{131}I and thyroid conditions, the magnitude of risk per unit dose, and the public-health question of how much risk was associated with ^{131}I in the population of children who were downwind from Hanford. Another potential way to address the public-health question is to compare the incidence of thyroid cancer or other thyroid conditions with the incidence in unexposed populations. However, comparisons with an external, general population are fraught with problems. Persons living in various geographic areas might vary in their baseline risk of thyroid diseases because of differences in dietary iodine intake and other unknown factors. Perhaps more important, the rates of detected disease are based on examinations and depend on the methods and criteria of the examinations; this produces screening effects that cannot be readily disentangled to make meaningful comparisons with disease-rates from other geographic regions that did not have comparable screening.

The HTDS investigators attempted to compare the number of thyroid cancers that they detected with the number expected in the general population. They reported that the observed number of thyroid cancers and the number expected in the general population were almost identical. To do that, they had to introduce a factor to account for their study group's having received a thorough thyroid screening, whereas the general population by and large has not received one. They chose a screening factor of 3, which had been reported in a 1985 monograph on radiation-induced thyroid cancer (NCRP, 1985). But that factor was based

on only indirect evidence: specifically, the prevalence of nodules found by screening in two studies was multiplied by 0.1 or 0.12 at various ages because a third study found that about 10–12% of nodules were malignant, and this result was compared with the incidence reported in a national survey, which proved to be one-third as high as the prevalence found in the two screening studies. That is a weak and questionable basis for choosing a multiplier of 3—there could have been unaccounted-for differences among the studies, and the screenings involved only palpation and not ultrasonography, as in the HTDS.

More recent studies, which were available but not cited by the HTDS investigators, have produced different values for a screening factor. For example, the study of atomic-bomb survivors, which at different times involved only palpation or palpation plus ultrasonography, produced a screening factor of 2.5 (Thompson and others, 1994). A study in Chicago with a sensitive screening technique produced screening factors of about 7 for thyroid cancer and 17 for thyroid nodules (Ron and others, 1992). The discrepancies in those values indicate that there is a great deal of uncertainty in the appropriate size of the screening factor, and the different values could allow one to conclude that those residing near Hanford had anywhere from a large deficit to a small excess of thyroid cancer. Hence, there is no unambiguous answer.

The HTDS Draft Final Report does not indicate any attempt to compare the HTDS thyroid-nodule prevalence with that found in unirradiated populations. Reported prevalence rates in unirradiated groups are available from about a dozen studies in the literature, so, in principle, it is possible to do, although again there would be a question about comparability with respect to screening intensity.

In summary, in the subcommittee's conclusions drawn from comparisons with general-population prevalence would probably have more uncertainty than those drawn from dose-response comparisons in the study population, so the HTDS

investigators rightly chose to emphasize the internal comparisons rather than general-population comparisons.

ANALYSES OF SOURCE OF PERSONAL-EXPOSURE INFORMATION

One major component of the determination of individual ^{131}I exposures was the milk-drinking habits of the study subjects. An attempt was made to interview a parent of each subject or other knowledgeable surrogate to obtain recollections of the milk consumption of the subject in childhood in terms of quantity and sources of milk at various ages. However, for 38% of the subjects it was not possible to interview a parent or surrogate, in which cases default assumptions were used in calculating thyroid dose. The defaults that the CIDER model used proved to result in considerable overestimates of the average dose derived from the reported milk consumption and sources. Specifically, the doses using default values were 40% higher than the average dose of those interviewed. For the critical group who were infants in 1945–1946, the discrepancy was even greater: the doses using default values were 77% higher than the average of interview-estimated ones.

A table showing mean doses by amount of milk consumption in a given geostratum. would be illuminating in indicating the degree to which dose variations were driven by milk consumption versus geographic location. That is important for understanding the degree to which the study's negative results might have occurred because of lack of reliability or validity in the reported milk-consumption rate. If a large fraction of the variation in dose is attributable to milk-consumption variation, the random-error component of the dose estimates is probably large, considering that Dwyer and others (1989) found a correlation of only 0.3 between contemporaneous reports and long-term recall of milk-drinking habits; this implies that one would not be likely to detect a dose-response relationship. A similar table giving mean doses by source of milk information (interview versus defaults) and geostratum, would also be informative.

Analyses that take into account the source of milk information are needed. The HTDS investigators did perform secondary analyses that used defaults for those without interviews on the basis of average reported quantity and sources of milk, and they indicated no association, but actual results were not presented. A useful analysis would examine associations using only those with interview information so as to yield results that minimize dose misclassification. Section 6 of this report describes the effect of milk-consumption measurement error on the statistical power of the study.

6

Statistical Power and Interpretation of the Study

Statistical power is discussed in section VII of the HTDS Draft Final Report and in additional material given to the subcommittee: appendix H of the HTDS protocol (May 1993), section II.G of the HTDS Pilot Study Final Report (January 1995), and a memorandum by Ken Kopecky (December 14, 1995). This review focuses primarily on thyroid malignancies as an example, although most of or all the points can be made about the other disease end points of interest.

FACTORS IN STATISTICAL POWER

Because the study results were essentially negative (that is, a finding of no increase in thyroid disease among those with higher estimated levels of ^{131}I exposure compared with those with lower estimated levels), a critical issue is how to interpret the negative findings correctly. Part of the process of interpretation is to subject the study to a series of reality checks. Were the data of sufficiently good quality? Do the underlying patterns of exposure and disease agree with or counter the negative association? Are the confidence intervals wide enough for the results to be compatible with other studies that have found an association between ^{131}I exposure and disease? If, for example, this negative study did a very good job of estimating thyroid-disease rates but found that milk-drinkers have higher rates of disease than non-milk-drinkers and that those who lived directly downwind of the site have higher

rates of disease than those in the more northern ("low-dose") counties, we might conclude that the pattern of thyroid morbidity fits the likely pattern of exposure and that the lack of a dose-response association could well be due to poor dose estimates. Similarly, a negative finding that is due to limitations in the assessment of thyroid-disease rates, because of either poor data collection or small numbers of subjects, does not support a conclusion that downwinders' disease patterns are unrelated to Hanford exposure patterns.

The power of a study of this type to detect a hypothesized increase in disease prevalence per unit dose (an absolute risk of 2.5% per Gy was used in the power calculations for thyroid carcinoma) depends largely on

- The size of the sample.
- The background prevalence of disease in the sample studied.
- The absence of biases that are caused by subject selection, reporting, or inaccurate detection of disease.
- The dose distribution in the sample.
- The adequacy of the dosimetry system in characterizing the dose of an individual.
- The independence of disease between individuals in the study, conditional on dose (for example, if there is no geographic clustering of disease or systematic geographic difference in disease rates due to other unknown factors).

SAMPLE SIZE AND ASSUMED BACKGROUND PREVALENCE OF DISEASE

The HTDS was successful enough in obtaining subjects who were willing to participate in the study that it essentially met its goals, so sample size is not a concern if the statistical-power assumptions and methods were appropriate. For the second point listed above, it was assumed in section VII of the HTDS Draft Final Report that the cumulative background prevalence of thyroid carcinoma either previously diagnosed or detected in screening would be 0.7% in females and 0.4% in males; these translate into 19 expected background cases (in the absence of a Hanford dose) in the cohort. Thus, if there is no increase in thyroid cancer due to the Hanford exposures, the 20 observed cases closely match the assumptions made in the statistical design. (For further discussion of the expected number of cases, see chapter 5 of the present report.)

EFFECT OF DOSIMETRY ERROR ON STATISTICAL POWER

Primary issues involved in determining whether the statistical power of the study was as expected are the dose distribution of the sample and the precision of the dosimetry system at the individual level. Distributional assumptions must be made in computing the power of the statistical tests or sample size for a specified power, and one can ask how robust the results were to these assumptions. The HEDR project acknowledged that parameter values used in its dose-reconstruction process were uncertain. Some of the uncertainties (those associated with release, dispersion, deposition, uptake modeling, and so on) were common to many individuals' dose estimates, whereas others (associated with food consumption, lifestyle, and so on) were individual-specific.

The HEDR project expresses parameter-value uncertainty with subjective probability distributions, which quantify the state of knowledge as judged by the HEDR analysts. The propagation of uncertainty through the HEDR models results

in a subjective probability distribution for each individual's dose estimate. The resulting distributions are very complicated, so, rather than an analytic characterization, a random sample (the 100 alternative realizations of the HEDR dose) produced numerically served as an approximate quantitative expression of the combined influence of the parameter-value uncertainties on the estimation of the individuals' doses.

The random sample was drawn so that the correlation among individuals' dose estimates due to common sources of uncertainty would be preserved. To summarize the dose distribution for each person, the median of the 100 dose realizations was calculated, and the median forms the single "dose estimate".

The effect of dosimetry errors on the expected or achieved statistical power of the HTDS is not mentioned in the protocol section on statistical power. Instead, in the power calculations, the distribution of the median dose estimate for each person is used as though it is equivalent to true dose.

It would be valid to ignore the dosimetry errors in the calculation of the statistical power for detecting nonzero parameter values in a linear dose-response model if both the following criteria hold:

- The average value of true dose for all subjects with the same estimated dose is equal to the estimated dose.
- Dose errors are independent from subject to subject, or at least any correlation between subjects' true doses (given estimated doses) is due to additive rather than multiplicative components of error.

Consider the second criterion. If all the doses were off by a constant unknown additive amount, then only the intercept terms, not the slope coefficients in the linear models, would be

affected by the correlated dose errors. However, for most shared sources of uncertainty, a multiplicative effect is likely. For example, if all the errors in the doses were due to uncertainties in the milk transfer coefficient appropriate for herds near Hanford, this would affect all doses multiplicatively; the estimated slope terms would retain the uncertainty, even in an infinitely large epidemiologic study.

VIOLATION OF THE BERKSON ERROR ASSUMPTION

If the first criterion is met, the statistical literature indicates, the estimation of linear dose-response models is essentially unaffected by independent dosimetry errors. In fact, an important measurement-error correction technique, known as "regression calibration", consists of the calculation of the average value of true dose, given estimated dose (see chapter 3 of Carroll and others, 1995), and the substitution of this average in the regression analysis. The first criterion, is sometimes called the Berkson model of measurement error.

Berkson errors arise when dose estimates are given as the average value of possible doses of a category of subjects who individually have a range of possible doses. The aim of the designers of the dosimetry system (the HEDR project) was evidently to provide the same average dose for all members of a particular category defined by "input data" (such as specific geographic location on a particular day with particular meteorologic conditions and specific age). Because the input data do not by themselves define the dose precisely, a Monte Carlo procedure was used in which all possible factors affecting a given individual's dose were considered to be random and 100 possible doses were drawn. Use of the average of possible individual doses as the dose estimate ideally results in Berkson error. However, because uncertainties in multiplicative factors that affect all doses simultaneously are admitted by the HEDR project (source terms, milk transfer coefficients, on so on) even under ideal

circumstances, violations of the second criterion are expected. These violations can be considerable.

As described in section VII of the Draft Final Report, the noncentrality parameter governing the power of the study is equal to

$$NCP = 1/2NB^2\sigma^2[\frac{1}{p_m(1-p_m)} + \frac{1}{p_f(1-p_f)}] \quad (1)$$

where

N = sample size (here 3,190),

B = assumed dose-response relationship (absolute risk, 2.5% Gy for thyroid malignancies), and

p_m and p_f = cumulative incidence of thyroid malignancies in males and females, respectively, in the population as a whole (assumed to be 0.4% and 0.7%).

The variance term ^{2} has to do with the variance of the dose distribution. If doses were observed without error, this term would be equal to the variance of the dose estimates. When doses are observed with error (whether Berkson or from any other model) but errors are independent or dependent because of purely additive components (that is, satisfying the first criterion), ^{2} is replaced in equation 1 with the variance of the average of true dose given estimated dose: ^{2} = Var (Avg (True dose/estimated dose)). (This follows as a consequence of the "regression-calibration" approach to measurement-error correction.) By definition, if one accepts the argument that the HEDR system produces Berkson errors and that the correlations are small, the value of ^{2} to use in

equation 1 is just the variance of the average dose for each individual. This is essentially what was done in the sample-size and power calculations for the HTDS, except that the median rather than the estimated mean dose was used. Dosimetry error always reduces study power, because it reduces the correlation between study outcomes and the exposure estimates, relative to the correlation that would be seen if true dose were known. However, in linear models of the probability of occurrence of disease and for dosimetry errors that satisfy the two conditions above, the slope parameter being estimated will remain statistically unbiased if the average dose estimates are used. Thus, the formula for the noncentrality parameter in equation 1 holds, except that the value of ^{2} being used is now the variance of the average dose estimates, that is, Var(Avg(True dose/estimated dose)), which is always less than the variance of the true exposures.

DOSE ERROR DUE TO UNCERTAINTIES IN INPUT DATA

One reason for doubt about the substitution of the variance of estimated doses for ^{2} is that the input data themselves are subject to obvious errors. The input data consist of such factors as location of residence (probably known fairly well) and milk-consumption habits in early childhood (undoubtedly known much less well; more will be stated about this below). If the fundamental input data are known with error, then in general Var (Avg (True dose/estimated dose)) will be overestimated by the dosimetry system. That occurs because the averaging process required to calculate Avg (True dose/input data) uses too few scenarios, and too little overlap is assumed between the scenarios that correspond to the distinct sets of reported input data.

The mean estimated dose in the HTDS Draft Final Report is 182 mGy with a variance equal to $(227\ \text{mGy})^2$. The sensitivity of the power of the study to the assumption that dose errors are of the Berkson type is approached as follows.

Suppose that the distribution of true dose is lognormal and that instead of a Berkson-error model we assume a classical-error model on the log scale so that

log(estimated dose) = log(true dose) + error. (2)

In this model, it is the estimated doses that are randomly distributed, multiplicatively, around the true doses. This model has often been considered potentially appropriate if input data are known with errors that are independent from subject to subject. The relevant aspect of the model here is that the average of true dose, Avg (True dose/estimated dose), derived from it has smaller variability, from person to person, than does the estimated dose itself. (The large estimated doses are reduced, and the small estimated doses increased.) The reduction of variance (of the average of true compared with estimated dose) is governed by the relative sizes of the variances of the last two terms—log (true dose) and error—in equation 2.

Assume, for example, that errors in log (estimated dose) have mean zero and standard deviation equal to 0.30 and are independent between subjects. This roughly corresponds to measurement error with a coefficient of variation of 30% (quite small compared with the variation seen in the 100 HEDR dose replications discussed in section VII, figure VIII.4). In this case, it can be shown that if the estimated dose distribution has a mean of 182 mGy and a variance of $(227\ mGY)^2$, the variance of Avg (True dose/estimated dose) would be equal to $(178\ MGy)^2$. Substituting that for ^{2} in equation 1 reduces the power from 96% to about 85%. Assuming larger errors in equation 2 has correspondingly larger effects on the analysis. Reduction of the power to below 60% (generally regarded as a study of low power) would occur when the standard error in equation 2 equaled 0.48, because this will reduce the Var(True dose/estimated dose) to about $(125\ mGy)^2$. Note that 0.48 is still quite small compared with the overall variability seen in the 100 estimates of HEDR doses. A value of

0.48 in equation 2 gives dose variations of about a factor of 2.8 compared with the even larger value of 4 seen in figure VIII.4 in the HTDS Draft Final Report.

To reiterate, the important thing about equation 2 is that the variance of the average true dose is smaller than the variance of the estimated doses. For example, if the CIDER program tended to overestimate the average true dose for all subjects by about 80%, the actual power of the test would again be 60% instead of the 96% claimed, because it would also correspond to $^2 = (126 \text{ mGy})^2$.

We conclude that if a substantial, but not overwhelming fraction of the variability of the HEDR individual dose estimates actually is due to non-Berkson error, as in equation 2, or if there is a substantial additional component of error due to uncertainties in input data, the power of the study likely was reduced to below levels that would usually be considered acceptable.

A worst-case scenario, in which all the error seen in figure VIII.4 of the Draft Final Report is due to independent errors in equation 2, would produce very low power to detect a positive dose-response relationship. For a number of reasons, however, it is considered unlikely that such a worse case actually applies. Given that the dosimetry system is based on extensive Monte Carlo calculations over many scenarios for each individual's input data, it seems reasonable to believe that some errors in the dose estimates do correspond to Berkson error. Also, the point is made in the later parts of the HTDS results section (section VIII) that a primary feature of the data is that two of the geostrata with the lowest estimated doses (Okanogan County and Ferry-Stevens Counties) actually had the highest rates of many of the thyroid diseases considered. Basic considerations of such factors as the prevailing wind directions would indicate that those counties should have had less ^{131}I deposition then the other counties in the study. Unless such basic assumptions in the dosimetry system are incorrect, it is difficult to believe that this is an artifact of dosimetry error.

EFFECT OF ERRORS IN ASSESSING CHILDHOOD MILK CONSUMPTION

The effect of errors in assessment of childhood milk consumption on the power of the HTDS to detect dose-response relationships depends on the fraction of variation, M, in between-person thyroid dose that is due to between-person variability in milk consumption. If R is the coefficient of correlation between reported milk consumption and true consumption, the sample size needed to maintain the same power, relative to a study of size N with no errors in reported consumption, can be approximated to a first order as $N/(1 - M + MR^2)$ (see appendix D). If, for example, half the variation in thyroid dose is due to variation in milk consumption and the correlation between true and reported milk consumption is 0.3, the sample size needed is 1.8N. Thus, about 80% more subjects are needed, in this example, to make up for the poor correlation between reported and true milk consumption. The HTDS Draft Final Report does not discuss errors in the milk-consumption estimates relative to the calculation of the HEDR doses, so we assume that no allowance for such errors has been made.

The HTDS Draft Final Report did investigate the effect of substituting defaults for estimated milk consumption (reference values, rather than each subject's individual estimate) in the HEDR model for thyroid dose estimation. The use of the reference values did not change the overall trends in the dose-response analyses, but more information about this analysis is needed. A comparison of the variance of the HEDR dose estimates using individual versus HEDR reference milk-consumption estimates would be helpful for two reasons. It would allow calculation of the statistical power to detect the hypothesized dose-response relationship in the analyses that used the reference diet—an important point not explicitly discussed. And, by allowing the estimation of M, it would partly address the extent to which the HTDS might have been over optimistic about the value of the retrospective reports of diet. In

particular, if the variance of the individual diet-based HEDR estimates is much larger than the variance of the reference diet-based estimates (that is, if M approaches 1), the power of the primary analysis (which used individual diet) could be very sensitive to low values of R for the correlation between estimated and true individual consumption of milk. (If M is close to 1 and R is 0.3, it would take a study size perhaps 10 times as large to obtain the power that knowing true consumption would yield.) But, if M is relatively small, then the power of the HTDS to find dose-response relationships is much less sensitive to assumptions about the accuracy of the retrospective surrogate reports of diet because other factors (such as location of residence) dominate the calculation of the estimated doses. Both power (of the reference diet analysis) and sensitivity (of the primary analysis) to error in the individual diet estimates should be discussed in future revisions of the Draft Final Report.

CORRELATED DOSE ERRORS

Correlations between individuals in dose errors also affect the power of the study to detect a dose-response relationship. For example, if the CIDER program tended to overestimate the average dose for all individuals by about 80%, the actual power of the study would be 60% instead of the 96% claimed, because it correspond to $^2 = (126\ \text{mGy})^2$. But if doses were consistently underestimated, the study power would increase. In general, highly correlated multiplicative errors lead to wider confidence intervals (and hence reduced power) for estimated slope terms in the linear model, inasmuch as allowance for the common uncertainties in dose need to be accounted for. Because the Monte Carlo procedures used by the HEDR project involved averaging over possible values of a number of parameters (source term, milk transfer coefficients, and so on) that are expected to affect doses multiplicatively, some analysis of the correlation between doses should have been performed as a part of the power calculations for the HTDS.

EFFECT OF GEOGRAPHIC VARIATION IN DISEASE RATES ON STATISTICAL POWER

A notable feature of the HTDS data is that there were indications of heterogeneity by geostratum for many of the thyroid-disease outcomes considered. Part of this heterogeneity was that the two low-dose geostrata often had higher rates of diseases than the other areas, whether or not dose was considered as a risk factor. That sort of heterogeneity of background rates of disease can have important effects on the power of studies of this type. Essentially, the issue is related to the last statistical "factor" noted earlier (see page 100): whether, conditional on dose, disease outcome is independent from individual to individual.

It is possible that important known or unknown covariates for thyroid-disease risk, not considered in the study, could lead to biases or loss of power if they tend to cluster by geostratum, distance from the Hanford facility, or otherwise in ways that affect estimated doses. To take an unlikely example, adult weight has been found in case-control studies (McTiernan and others, 1987; Preston-Martin and others, 1993) to play an important role in thyroid-cancer risk, with the heaviest subjects in one study (Goodman and others, 1992) having up to 5 times the risk as the lightest. If for any reason the average weight of subjects differed substantially by geostratum or distance from Hanford, this could make the estimation of a dose-response relationship quite difficult. Similarly, thyroid-cancer rates have been noted to differ by ethnicity, with an especially high rate among Filipino women in Hawaii (Kolonel, 1985). The Hanford study population is ethnically homogeneous, but the evidence, for many of the thyroid diseases or abnormalities, that risk differs by geostratum raises the question of whether clustering of important unconsidered factors could have reduced the power of the study by violating the independence assumption.

7

Comparison with Other Studies

CONCORDANCE WITH THE UTAH STUDY OF THYROID DISEASE FROM NEVADA TEST SITE FALLOUT

Section IX of the Draft Final Report describes in some detail whether the HTDS study confirms the results of the Utah study of exposure to ^{131}I from the NTS (Kerber and others, 1993). With a generally negative dose-response relationship, especially for thyroid carcinoma, the HTDS cannot be regarded as confirming the Utah findings of increased risk of thyroid neoplasia. However, another aspect of the comparison of the two studies is not fully treated: the degree to which the results of the HTDS directly contradict those of the Utah study. One approach to answering that question is to assess the degree to which confidence intervals of risk estimates from the two studies overlap. Even if one study's results are positive and another is negative, it does not mean that they necessarily are irreconcilable. If the positive study is barely positive ($p \sim 0.05$) and the negative study has wide confidence intervals, there might be no fundamental disagreement.

For the HTDS analysis of thyroid carcinoma, the estimate and the confidence interval for the linear slope term for thyroid carcinoma are not reported, because the maximum-likelihood estimates failed to converge. We can, however, work backward from other information in the report to estimate the standard error of the slope term. The attained power to detect a

slope term of 2.5% per Gy is stated to be 0.96 (section VIII), so we will have (at least approximately)

0.025/(standard error of B) = $(Z_{1-0.05}-Z_{1-0.96}) = 3.396$.

Therefore, the standard error of slope term B must have been 0.007 per Gy. To approximate the value of the estimate of B, we note that a logistic model did converge and gave an estimated slope that was about 1 standard error below zero. We can assume roughly that the linear slope estimate would also have been about 1 standard error less than zero, or about -0.007 per Gy. If (as hypothesized) males had a background risk of 0.004, the upper confidence limit for the risk at 1 Gy is -0.007 + 2(0.007) = 0.007, so the upper limit of the excess relative risk (ERR) at 1 Gy is 1.75 for males. For females, assuming a background of 0.007, the ERR at 1 Gy is 1, so the average of the two is ERR = 1.375 per Gy. For the Utah study, the estimate was 7.7 with a lower 95% confidence limit of 0.74 per Gy. It seems, then, that the confidence intervals for the risk of thyroid cancer overlap to some degree. Moreover, on the basis of the considerations above, it is evident that the confidence intervals for the HTDS in fact depend on dosimetry-error assumptions; if the pure Berkson model of errors in the dosimetry does not hold, the confidence intervals for the HTDS could be considerably wider. Thus, there does not appear to be a fundamental incompatibility between the two studies.

CONCORDANCE WITH STUDIES OF EXTERNAL RADIATION TO THE THYROID

It may be an oversimplification to say that the HTDS, because it found no significant dose-response relationship for any disease end point, is in direct contradiction with the cohort studies of external radiation exposure and risk of thyroid cancer. Of the five cohort studies of external radiation and thyroid cancer that were analyzed by Ron and others (1995), one yielded estimated dose-response relationships considerably stronger than the others.

To combine the results of the five cohort studies, Ron and others used a random-effects model. The dose-response relationship was allowed to vary from study to study, and the average dose-response relationship for a hypothetical population of studies was estimated. The average estimate was equivalent to an ERR of 7.7 times the age-specific baseline per Gy with a confidence interval of 2.1–28.7 times the baseline. As described above, the HTDS is probably consistent with an upper ERR of about 1.4 per Gy, which is not statistically compatible with the estimate for external radiation. It is not known whether the two estimates could be statistically compatible if uncertainties in dosimetry were factored into the confidence interval for the HTDS.

COMPARISON WITH CHERNOBYL STUDIES

The effectiveness of ^{131}I in causing thyroid cancer has been shown by the Chernobyl experience. The first increases, reported in 1992, of thyroid cancer attributed to the accident were challenged as possibly the result of intensive screening. More recently (Astakhova and others, 1998), however, a case-control study in Belarus has found highly significant differences between cases and controls in estimated ^{131} I dose to the thyroid, even when controls with similar presenting complaints or screening circumstances were selected. But the durations of exposures were shorter for Chernobyl than for the Hanford downwinders, and the doses were higher, so the dose-rate issue still is unresolved with respect to the epidemiologic data. Furthermore, the dose reconstruction for the study in Belarus was based on actual measurements of ground deposition of ^{131}I and cesium-137, a data bank of 1986 thyroid-radiation measurements, and interviews and questionnaires. Therefore, doses were probably better estimated for individuals in the Belarus study than in the HTDS.

8

Communication of Study Results

There are many factors to evaluate in the communication planning and practice related to the release of the HTDS Draft Final Report. Some aspects of the communication strategies were well carried out, particularly the ones that kept the public informed about the study protocols, design, and progress over the years. However, in the fall of 1998, problems—some controllable and some not—arose that had substantial and unfortunate effects on the communication efforts that finally were made in late January 1999 for the release of the HTDS findings. The release of the Draft Final Report led to unhappiness and dismay among some citizens in the Hanford area, not only because of the main message, but also because of how the message was delivered.

BACKGROUND

The communication of risk information to the public is an important issue that has been addressed by many individuals and groups. Numerous journal articles, manuals (Swanson and others, June 1991), and reports—including one by the National Research Council (1989)—have discussed how to convey information about health and environmental risks to the public. But there is no sure prescription to follow for providing such information effectively. One piece of advice that appears in most risk-communication publications is to ensure that the information

source is credible and trustworthy in the minds of the public to which it is communicating.

Risk communication as a field has evolved over the last 15–20 years from a simple linear top-down communication model (experts translated technical information and dispensed it to the public) to a more sophisticated series of models that view risk communication as a "tangled web" of interactions that move in many directions and involve many players—local, state, and federal governments; special-interest groups; citizen groups; industries; unions; and so on (Krimsky and Plough, 1988).

The communication that occurred about the HTDS is as tangled a set of risk-communication webs as one comes across. It includes not only a federal agency and a private contractor, but also health agencies in three state governments and representatives of nine American Indian nations, numerous citizen groups in the region, national and regional journalists, a class-action lawsuit involving many litigants, various consultants and potential expert witnesses, and a number of private individuals in the region who have suffered or whose family members have suffered from some type of thyroid disease. The public and private messages traded back and forth by these groups and individuals over the years have all shaped the Hanford and HTDS communication process. For example, it is important to remember that, according to citizen comments reported in the mass media and elsewhere, many of the citizens in the area had developed a distrust of government sources, particularly the Department of Energy (DOE) and its precursor agencies. When CDC began to evaluate the thyroid-disease situation at Hanford, the distrust was already in place; although not applied directly to CDC at the time, it became an important factor when the HTDS Draft Final Report was released. From citizen comments, it appears that the Fred Hutchinson Cancer Research Center in Seattle enjoyed greater public trust than the federal agencies while carrying out the HTDS.

ORIGINAL HTDS COMMUNICATION PLAN

The HTDS was conducted over a 9-year period, and it is not possible to evaluate all the written and oral communication with the public that occurred about it, but apparently a number of potentially effective communication efforts were made. As described in section X of the HTDS Draft Final Report, public meetings on the project began in 1990 and continued throughout the study, written brochures and fact sheets were developed, newsletters were sent to more than 9,000 people and organizations, and a toll-free 800 line and a Web site were available. In March 1991, the first meeting of the federally appointed public HTDS Advisory Committee was held, and this committee and several other groups approved the study protocol. The advisory committee continued to meet throughout the life of the study. Special arrangements were made to keep study participants advised of the results of their clinical evaluations of thyroid disease. The open communication seems to have continued almost up to the end of the study, and no one who provided information to the present subcommittee during its public meeting in Spokane or otherwise criticized those communication efforts.

Given the earlier history of less than openness with the public in the Hanford region on the part of DOE and its precursor agencies, this plan for open communication was enlightened and promising. So were the decision to establish a citizen advisory group for the study and the apparent cooperation offered to various other citizen groups in the region, including the Hanford Health Information Network and the Hanford Health Effects Subcommittee (HHES). All those efforts seemed to help to build trust and credibility for the study and its investigators and for CDC.

Section X of the HTDS Draft Final Report outlined a good communication plan to deliver the final information about the report that might have worked if it had been put into operation with its release in March 1999. Especially admirable was the concern shown for translating the technical information in the Draft Final

Report into an understandable booklet for the public. The use of focus groups to look at various presentation strategies for the booklet was a fine approach and resulted in a readable and understandable public summary. (More will be said about the public summary later.)

Considering the many audiences that had to be informed about the study findings, the communication plan described in section X was fairly complex and involved several types of briefings, with a suggestion that some of these be conducted via satellite connections to remote broadcast sites throughout the region. Although expensive, such a plan would have probably worked far better than the telephone briefings that were eventually used when the Draft Final Report was released. The original communication plan was approved by the HTDS Advisory Committee and widely disseminated in the HTDS newsletter. According to Scott Davis, one of the principal HTDS investigators, there was never any intent to release to the public the draft report that was delivered to CDC at the end of September 1998. That draft report was supposed to go through internal CDC and National Research Council closed peer review; comments were to go to the HTDS researchers, who would then make needed changes and release a peer-reviewed final report in March 1999.

However, three major factors interacted to bring about the early release the Draft Final Report, which contributed to the extensive communication problems encountered: public pressure to get the document out to the public and concern that CDC's internal review would alter the original findings of the HTDS investigators, the National Research Council's desire for open peer review of the draft report, and a subpoena from one party in a lawsuit that sought immediate release of the Draft Final Report.

Concerning the first factor, in early October 1998, the director of CDC's National Center for Environmental Health received numerous written requests for immediate release of the Draft Final Report. According to CDC, because the agency had received the Draft Final Report only on September 30, the requests

indicated that the correspondents expected the report to be made available to the public unchanged.

As to the second factor, a communication in late October to CDC from the National Research Council said that this organization felt that the credibility of its review would be compromised if the HTDS report were not publicly available when the review process began. Given the many audiences, including study participants who needed to be informed about the study findings, given the extensive communication plan already developed, and wanting to preserve the credibility of the review process, CDC officials decided to release the Draft Final Report before the Research Council's open review. On November 12, they met with the HTDS investigators and decided to release the report on January 28, 1999.

The third factor added resolve to that decision. According to CDC, during the week of November 16, 1998, the HTDS investigators received a subpoena that called for delivery of the Draft Final Report to the plaintiffs' lawyers within 30 days. The delivery date of the report had already been moved up, and the plaintiffs' attorneys, with consent of the court, indicated that they could wait until the January release. One of the problems preventing an even earlier release was the need to develop documents about the report that the public could understand (Davis, 1999b).

Both public demands to CDC to release the report and not change it were exemplified by the minutes of an HHES meeting held on December 10—11, 1998, in which a CDC official explained that the agency had heard from many people "who expressed a real interest in seeing if we could move the date of the release of this report up to make it public" (HHES, 1998). Concerning fears that changes would be made in the draft by CDC's internal reviewers before the report was released, one person said, "if there are changes made between what Fred Hutchinson delivers and what comes out the door at CDC, I'm hoping that you have heard from this subcommittee clearly that we

want to know what those changes were and the rationale for those changes" (Hanford Health Effects Subcommittee, 1999, p.39–40).

Concern also was expressed at the meeting about plans to release the report, in particular, how HHES and other citizen groups would be briefed and the fact that the HTDS investigators were submitting an article on the study results to a scientific journal before citizens knew the results.

COMMUNICATION ISSUES IN THE WRITTEN MATERIALS

Besides the Draft Final Report itself, a number of written pieces were developed for the public, including the "Summary Final Report of the Hanford Thyroid Disease Study" and several HTDS newsletters. One featured a summary of the study results, another presented information on thyroid disease and how it was diagnosed, and a third included questions and answers about radiation and thyroid disease. There were also at least seven fact sheets from CDC on various subjects related to the HTDS. Most of these were included in the briefing kits for the media and public when the report was released.

To evaluate the written materials provided by the HTDS investigators and CDC, a number of different factors must be considered. Among these are accuracy, appropriateness of material, and readability or ease of understanding.

Accuracy

From a scientific perspective, the results section of the Draft Final Report provided details on linear dose-response analyses conducted on 13 types of thyroid disease or abnormality and on a number of thyroid-related laboratory tests. Other analyses were conducted, when possible, with alternative definitions of the disease being analyzed and alternative dose-response functions. None of the numerous tests showed a statistically significant increase with dose.

The resulting database on the absence of dose-response relationships formed the basis for statements made in the report that no evidence was found of a statistically significant increase in effects associated with increased dose of ^{131}I. Within the reported framework of the data used and the analyses conducted, that is an accurate representation of the results given in the Draft Final Report.

Appropriateness

Although the above statement is scientifically accurate, given the state of a draft not reviewed by outside scientists and a number of uncertainties that were already apparent to the HTDS researchers, some overstatements were made in the public summary, HTDS newsletters, the news release, and the executive summary of the Draft Final Report. In addition, there was little or no mention in any of these documents of the uncertainty issues involved in the study. Such uncertainties, already described in the present report and others, included the statistical power of the analyses, possible errors in the dosimetry, and the reliability of some information in the computer-assisted telephone interviews related to possibly faulty recall about milk sources and amounts. At the time of the Draft Final Report's release, the HTDS investigators were trying to run an uncertainty analysis but were not succeeding in its execution; this was not mentioned in the written materials.

Despite those problems, the results of the HTDS were presented with unqualified certainty, and at the time of and after the release of the report some statements attributed to the HTDS investigators appear to have overstated the certainty of the results in the Draft Final Report (emphasis added):

> Thus, given that the HTDS had adequate statistical power to detect reasonably small effects, and the rigor of the study design, these results provide rather strong evidence that exposures at these levels to ^{131}I do not increase

> the risk of thyroid disease or hyperparathyroidism. These results should consequently provide a *substantial degree of reassurance* to the population exposed to Hanford radiation that the exposures are not likely to have affected their thyroid or parathyroid health
>
> [Davis & Kopecky, 1998, p. 18].
>
> Findings of the Hanford Thyroid Disease Study are *clear and unequivocal*
>
> [Davis and others, 1999].
>
> This was a *very powerful study* because it included a large number of people estimated to have a wide range of exposures to ^{131}I
>
> [Davis, 1999a].
>
> The study had *sufficient statistical power* to detect increases in thyroid disease risk that were predicted based on studies in other populations
>
> [CDC, 1999a].
>
> The design and successful completion of the study ensured a *very high probability of detecting relationships* between Hanford radiation dose and diseases under study if such relationships exist. The study *was very powerful* because ...
>
> [HTDS Newsletter, 1999].

The subcommittee believes that such statements were ill-advised at a time when the Draft Final Report had not yet been subject to external peer review. Given the many questions raised

about dosimetry and other issues and the problems in running the uncertainty analysis, the subcommittee feels that statements indicating such certainty should have been modified to take uncertainties into account and the uncertainties should have been listed and explained. Several paragraphs and perhaps a list of uncertainty issues should have been in the executive summary and all public documents related to the HTDS.

The subcommittee recognizes that including such uncertainty information would probably have diluted the strength of the investigators' message; however, such caveats are critical to increasing public acceptance of the results of the report. Omitting them left the investigators open to the charge that they had emphasized negative results.

On a related matter, the printed HTDS public summary was titled "Summary Final Report of the Hanford Thyroid Disease Study" (FHCRC 1999b), which was misleading, inasmuch as it was still a draft. From the cover of the printed summary, a reader would not have been able to tell that the report was a draft. Nor did the text of the public summary make it clear that the report was still under review and that some findings might be changed. The text explained that there had been public and scientific review of the study, but it implied that review was finished (FHCRC 1999b, p. 8). Those items gave too much certainty to the results and the study itself.

In contrast, the January 1999 HTDS newsletter included information in two places about the final stages of peer review by both the Research Council subcommittee and a journal, and it called the report the "Draft Final Report to the CDC". Later versions of the public summary given out by CDC were also stamped "Draft".

The subcommittee recommends that all uncertainty issues be clearly noted and explained in the final report and all public documents related to it.

Contextual Information

Both the main public documents and the executive summary of the Draft Final Report lack comparisons with findings in other reports on similar subjects, such as the Institute of Medicine's report (NAS/IOM, 1999) on the nationwide fallout study by the National Cancer Institute (NCI, 1997) and the NTS report (Kerber and others, 1993). It should be noted, however, that a CDC fact sheet, "What We Know from Other Studies of Environment Radiation Exposure and Thyroid Disease", was available. How widely it was disseminated is unknown. The subcommittee believes that not having at least a summary of this contextual information in the public summary and the Draft Final Report presented a problem because readers did not have any background information with which to compare the HTDS results if they did not have the CDC fact sheet. In risk issues, many readers and journalists need to know more than just the event at hand—in this instance, the findings of only this one report. They need to see the long-term issue in context so that they can understand the variety of findings and judge for themselves the validity of the current study against the others. The subcommittee recommends that contextual information be included in both the executive summary and the body of the final report and the public summary.

Readability

The public summary, fact sheets, newsletters, and Website information about the HTDS Draft Final Report were readable and relatively easily understood. However, the executive summary of the Draft Final Report was far from that. The language was technical and often highly complex, and it did not need to be so in most places. Scientists will not be the only people reading the executive summary, and it should be understandable for a number of educated groups (such as nonscientist government officials, lawyers, journalists, and social scientists), even though it need not be as simplified as the public summary.

There were a number of readability problems in the executive summary. For example, in the first part (pages 3–7), it was difficult to keep track of the many numbers related to how many people were in various parts of the sample. On page 4, in the fourth paragraph, the way the numbers were presented was confusing: 3,865 "potential participants ... agreed on either the first or second attempt. The remaining 3,565 ... agreed to participate...." Those sentences should be rewritten. Throughout this section, charts would help to clarify which numbers apply where and help the reader to follow the discussion. The material on the computer-assisted telephone interviews and on how people provided names of relatives to be interviewed (pages 4–5) also needs to be explained better.

Throughout the executive summary, abbreviations and uncommon words were unexplained or were explained after they had already been used. For example, finding "CIDER", "Exp.-IPI", and "realizations" on page 5 of the executive summary would probably confuse a reader. Those and other terms need to be briefly explained in the text or in footnotes.

The subcommittee recommends that an effort be made to remove excess technical language and to use consistent terms, particularly for types of thyroid diseases, in the executive summary of the final report. More charts should be used in the executive summary to provide visual aids to help understand the information.

In addition, the entire final report should be edited and should include a glossary of abbreviations and technical terms used.

Written Materials from CDC, May 1999

After the adverse response to the release of the Draft Final Report by members of interested citizen groups and some other citizens in the region, as shown in letters to the editors in several newspapers and to CDC officials, CDC prepared some new materials that were used at two public meetings, in Spokane and Seattle, to discuss the report in May 1999. In the "Summary of the

Preliminary Results", CDC (1999b) was more cautious about interpreting the findings:

> These preliminary results do not mean that people living in the Hanford area during the 1940s and 1950s were not exposed to ^{131}I and other radionuclides, or that these exposures had no effect on the health of people living in the Hanford area. Although no link between estimated ^{131}I and amount of thyroid disease was identified by the HTDS in the study population, the study results do not prove that a link between ^{131}I and thyroid disease does not exist. There may be individuals in the overall population who were exposed to Hanford radiation and did develop thyroid disease because of their exposure.

However, in backing away from the certainty that was a main theme of the January release, CDC might have gone too far. In using "preliminary results", the agency seemed unwilling to acknowledge that the study had reached the final-report stage. Later in the document, CDC referred to the 'initial study results' provided in the Draft Final Report. Although the peer-review process had not been completed, it is clear that the Draft Final Report of this 9-year, $18 million study had progressed well beyond the point of preliminary analyses. Such terms are normally reserved for periodic progress reports. Both underinterpreting and overinterpreting the results of this major study are problematic.

Updating the Communication Section of the HTDS Draft Final Report

The subcommittee recommends that the communication section of the HTDS report be updated to reflect the development

of a new plan to release the final version of the report and to provide some history of the communication efforts made for the Draft Final Report. The new plan should include such issues as whether the whole study or only the changes (if any) will be released and how uncertainty will be discussed. If the changes made to the Draft Final Report are minor, only minor planning is needed. However, given the public dismay at the release of the Draft Final Report, the subcommittee believes that major communication planning will be required to ensure that the integrity of the study and the investigators is maintained in the eyes of both the public and the media. Some plans will also be needed for the eventual publication of the article submitted to a scientific journal, if it is accepted, because its publication could attract more media attention. Additional requirements for a new communication plan will be discussed below.

RELEASE OF THE DRAFT FINAL REPORT

With hindsight, one can often see why a reasonably well-planned risk-communication plan was not successful. However, during the planning of a risk-communication effort, it is often difficult to evaluate how members of the public will respond to messages or whether they will even pay attention to them. There was little worry that the messages in the HTDS Draft Final Report would reach an interested population, including national and regional journalists. However, as described earlier, three major unplanned factors led to an early release of the report. In addition, at the time of the release, another unplanned factor—a leak to the *New York Times* about the report—put more stress on the release situation. To evaluate the outcomes of the various briefings that occurred with the release of the report, it is important to look at some different aspects, if only briefly: the planning process, the need for an information blackout, the briefings themselves, the leak to the *Times*, and the selection and effect of the main messages presented at the briefings.

Planning Process

To release a report of great interest and magnitude, such as the HTDS report, is not a simple task, and neither the investigators nor the CDC appeared to take it lightly. Plans described in section X of the Draft Final Report were ambitious and generally well designed. With the early release of the report, efforts had to be speeded up, and this possibly led to some problems because many audiences had to be dealt with: HTDS participants; the various advisory and citizen groups in the region; state public-health officials and state, county, and local government leaders in three states; tribal officials; Washington, DC, officials, including those at the Department of Health and Human Services (DHHS) and congressional delegations from the three states involved; and, of course, the mass media. In addition, the printed materials for the public had to be prepared as did material that would appear on the HTDS Web site with the release of the report. Mailings with study results had to be sent to hundreds of people, including the study participants. That is a great deal to accomplish in the 3 months between the decision to release the report early and its official release date.

Nevertheless, CDC had a well-planned "rollout" schedule, listing all the tasks, who would do them, and the necessary deadlines. Complicating the planning and material-development process was the need to procure many clearances in CDC and DHHS that took about 2 weeks, a "fast-track" timeframe. Some concern has been expressed that as some of the draft public materials went through the clearance process, some of the qualifiers on the findings were dropped and the message became more positive. That is often the case when many messages have to be cleared through channels, and it might have occurred here, but no evidence to that effect was seen.

Need for an Information Blackout

Part of the planning appeared to include an information blackout of the HTDS results until official release, of the report.

Supposedly, the blackout would let the study participants and the public know the study results at about the same time. That was somewhat unrealistic, given the number of groups to be briefed. The more briefings scheduled, particularly if some are a day or so in advance of others, the greater chance that information will be leaked—as it was in this case. In addition, the blackout had an unhappy effect on the citizen groups, including the HTDS Advisory Committee, which had been kept informed about the study's progress over the years. Members had come to expect to be updated about what was happening in the study and became upset about not being told the main findings earlier than the day of the public release.

In retrospect, there were reasons to keep the findings of the study confidential, but perhaps the most important—need to communicate first and foremost with the public—was downplayed. Trying to brief so many official parties before the public created a substantial opportunity for information leaks. This subcommittee believes that trying to maintain an information blackout, given the number of briefings needed, was problematic and unrealistic.

Briefings

According to the schedule provided by CDC, briefings about the report were to start in Washington, DC on January 22 and 26 with officials of CDC, DHHS, DOE, the National Institute for Occupational Safety and Health, and the Agency for Toxic Substances and Disease Registry. A briefing for congressional delegations was scheduled for January 27. On the main release day, January 28, the briefings moved to the state of Washington. Two morning conference-call briefings were scheduled for state health officers and the Northwest Tribal Nations and Indian Health Service. On the 28th, at 1 pm, four citizen groups, including the HTDS Advisory Committee, were to be briefed by conference call. The media briefing was to occur at 3 pm, and the public meeting on the report was scheduled for 7 pm in Richland, WA. Material concerning the study was to be posted on the Web site at 3 pm.

It was an ambitious schedule, and it represented a compromise from the earlier discussion in the Draft Final Report of using satellite connections for remote meetings rather than conference calls. One participant in the calls felt that they did not work well, because they were far too impersonal and unwieldy. Participants did not know who was on the other end; they could not show graphs or other illustrations; they could only deliver an abbreviated version of information that would be presented at the media briefing and take questions.

There also was frustration for the people on the receiving end of the conference calls, according to information given to this subcommittee. They briefly heard a message that they had not expected, had few details about the study, had nothing in writing, and could only ask general questions. In addition, because they had no written materials, they could not respond to questions being directed at them about the report by journalists on January 28. Another problem in retrospect was that the briefings—even those for the media or the public—were not taped, so they could not be transcribed later for interested parties. Several people in the region concluded that CDC did not take the briefings seriously enough to record them for review.

Given that there were no transcripts, it is difficult to evaluate how well the briefings were done. One reporter at both the media briefing and the public meeting was surprised by "how absolutely confident the Hutch people were." She pointed out that subtleties and uncertainties were not discussed, nor were any problems with statistical power. She noted that scientists usually are not that positive about their studies and often make "conditional statements particularly when a study is still a draft and hasn't undergone peer review" (Steele, 1999). However, another reporter who attended the media briefing said that even if the uncertainties in the study had been stressed, the media probably would not have emphasized them. She noted that the press "wouldn't have dwelt on the uncertainties", because the media, particularly the broadcast media, do not get into all the technical

details, and she said that they would report only that "the bottom line is this. That's the way the media operate" (Cary, 1999).

Effect of the Leak to the New York Times

After the congressional briefing on January 27, the HTDS results were leaked to a *Times* reporter, who quickly wrote an article about them. The story was picked up by the Associated Press and was on the wire to be seen by reporters in the Hanford area about 8 pm PST that night. It was early enough, said one reporter, that she was able to add some local reaction about the findings to a story that she had been writing about the release of the report the next day. Her newspaper also ran the *New York Times* story on the morning of January 28, before both the media briefing and the public meeting (Cary, 1999).

Again, an uncontrolled situation had changed a carefully planned risk-communication strategy. The leaked story sent reporters and HTDS and CDC officials scrambling. CDC media officials starting faxing materials about the study results to reporters at 6 am EST on January 28, not waiting for the 3 pm media briefing. They also began putting all the planned information about the report on the Web site at 3 am EST.

It is hard to evaluate the impact of the article with the leaked information on the planned risk-communication process, but several members of citizen groups said that they were upset because reporters were calling them for comment on the morning of January 28 and they had not yet been briefed. Even after they were briefed, they still had no written data and had not read anything official about the study to which to respond. They felt that they were put into an awkward position. There also has been some supposition that reading the results of the study in that day's newspapers made the people who attended the public meeting in the evening more angry than they would have been otherwise. It is hard to know whether that is valid. However, as one reporter said, the *Times* story set the tone for most of the media coverage that followed, and it was headlined "No Radiation Effect Found at

Northwest Nuclear Site" (Wald, 1999). Many who presented information to this subcommittee during its public meeting in June in Spokane complained about the media coverage of the study and its implied dismissal of thyroid problems at Hanford. That point is exemplified by a headline found in the *Salt Lake Tribune*, topping an AP story on the study: "Study Disputes Hanford Poisoned People" (Associated Press, 1999). For a number of people in the Hanford region, such a conclusion was unacceptable, and they blamed the media position on the overpositive and strong message provided in the Draft Final Report and the various briefings.

The Message and its Effect

In evaluating the risk-communication process and the activities that occurred around the HTDS Draft Final Report, it is hard not to question whether the public dismay with the release of the report would still have come about if the message had been different. The main message—no link between radiation exposure at Hanford and prevalence of thyroid disease—was not expected by concerned members of the public in the region. Given the findings of the NTS, Chernobyl, and other studies and the documented radiation releases from Hanford, a positive association was expected, according to interviews with local journalists and some concerned citizens.

Delivering unpopular risk messages is itself risky. In many instances, it has to be done delicately, with great thought about how it will affect an audience expecting an opposite result. Varied audience responses have to be forecasted and planned for. Sensitivity to audience health concerns and fears needs to be shown. In this particular situation, with an audience very concerned about perceived high rates of thyroid disease in the population—an audience that had been reported to have little trust in government agencies—great care should have been taken to deliver the results of the HTDS sensitively and tactfully. Implications for individuals and families that have suffered from

thyroid disease should have been not only explained, but also highlighted.

Perhaps knowing that they faced a difficult task, the HTDS investigators felt that they had to deliver as strong and positive a message as they could about their findings, indicating to people in the region that they should feel relieved that no link had been found. But that was not a sensitive way to proceed, given the audience. One important error was to emphasize the statistical group effect and not the outcome for individuals. It was only in response to questions at the briefings that they acknowledged that their findings did not negate the suffering of people in the region from thyroid disease. Later, they explained that they had left the uncertainties in the study undiscussed during the briefings and in the written public material because the focus groups that they had worked with on the public summary during the fall of 1998 had told them that anything technical was not appropriate for the public materials.

CDC's role in the message selection and delivery is more complicated. People providing information to this subcommittee in Spokane questioned why CDC had not intervened to counter the overpositive message about the study given by the HTDS investigators in the briefings and the written materials. They said that the investigators were contractors and that CDC was ultimately responsible for what was said about the study. They charged that CDC had done a disservice to the people of the region. Clearly, this is an important and complex issue. It involves agency-contractor responsibilities and relationships, academic freedom, and responsibilities to the public. It is even more complicated if one remembers that in December 1998, HHES members and other citizens urged CDC not to alter the report as it came from the investigators and to release it as it was. Those concerns helped to put CDC officials between a rock and a hard place. If they asked the HTDS group to soften the tone of the findings, they could be accused of altering the investigators' report. It is apparent that CDC officials would have been criticized

by various groups no matter what actions they took, because of the numerous audience expectations about the study results.

CDC officials said that the main message was decided by collaboration between them and the HTDS investigators on the basis of numerous discussions. Several CDC officials noted that they had concerns that some of the messages were too strong, such as "this was a powerful study." Despite their concerns, after struggling with some of the language in the report, they decided to leave it as drafted by the HTDS investigators because of public pressure not to alter the report.

The subcommittee recognizes and supports CDC's sensitivity to citizens' concerns and the needs of academic freedom for investigators, but it believes that there was a middle ground: both the HTDS investigators and CDC officials should have expressed their own views and interpretations about the Draft Final Report at the briefings and in the public documents. Although consensus might have been preferable, the differing interpretations should have been presented to the media and the public. That is preferable to having one point of view dominate the other—regardless of which side dominates—and then backtracking to change or soften a message. Such advice does not agree with some generalized risk-communication guidelines, but such guidelines must be adapted to specific situations. In this case, many members of the public and the journalists in the Hanford region were actively engaged in the issues and educated about them. They were not going to accept a simplified approach and message, particularly if it disagreed with their own experiences and points of view. Rather than presenting a black and white picture of the results with a positive spin, the HTDS and CDC personnel should have emphasized the shades of gray.

Despite its sensitivity to problems with the language in the report and concerns over what to do about it, CDC itself showed insensitivity to people and families with thyroid disease in the region when it announced at the public meeting on January 28 that it would recommend a change in plans for medical monitoring.

No matter what reasons officials gave—including a report by the Institute of Medicine questioning the value of medical monitoring—the public linked this action to the announced results of the HTDS. And although CDC took great pains to point out to the public and the media that the HTDS report was a draft and would be subject to peer review and public review, the agency appeared to be basing policy decisions on it. Even if the decision regarding medical monitoring was correct, announcing it at the same time that the HTDS Draft Final Report was released was a mistake and hurt CDC's credibility.

Recommendations about Releasing the Final Report

The subcommittee recommends that CDC continue its open-communication policy on the HTDS and improve on it for release of the final report. It applauds the development of materials for the public—such as the newsletters, the background fact sheets, and the Web site—and recommends that it be continued. It is important to remember that those efforts were well received in the community, and they should not be overshadowed by the problems encountered with the release of the Draft Final Report.

In writing and releasing the final report and its public summary, steps must be taken to explain alternative interpretations of the data and to ensure that findings are presented in an evenhanded method that does not overemphasize one point of view. Efforts must be made both in the report and its accompanying public documents to explain the implications of the findings for individuals and families sensitively, indicating, for example, that a statistical study does not necessarily negate the existence of thyroid disease in this population and explaining why that is so. The messages in the final report must be developed with sensitivity to audience health experiences, concerns, and fears.

Any substantial changes made from the Draft Final Report should be clearly outlined and explained, including why they were made. Remaining uncertainties must be acknowledged and explained, along with their implications and effect on the final

conclusions. CDC and HTDS personnel should work together on the wording and presentation of any public messages, presenting alternative interpretations of data and conclusions as needed. Given the controversy that already exists over this report, presenting only one viewpoint will lead to more public distrust.

A new communication plan should be developed for the final report. It must take into account and acknowledge the problems encountered with the release of the Draft Final Report and include dealing with possible lowered public trust in the HTDS investigators and CDC in the Hanford region. Because of serious problems encountered in trying to maintain an information blackout, such efforts should be minimized. For the final report, multiple briefings should be abandoned, and there should only be an early briefing for CDC and DHHS officials followed quickly, if possible given political realities, by one large briefing for all other parties, using satellite transmission or other advanced technology to link groups in various locations. In particular, citizen groups that have participated in the study process over the years should not be kept out of the information flow until the last minute. All media and public briefings should be videotaped to provide a record of the proceedings. Journalists should receive copies of the final report several days in advance of its official release after agreeing not to write stories about it until the release. That practice, known as embargo, is widely used and, particularly on complicated subjects, allows journalists time to study a report and develop thoughtful articles.

Given the investment of time and effort by the people who participated in the HTDS, they should be randomly surveyed as to their satisfaction with the communication of the draft study findings and their own dose results, so that communication to this group can be improved for the final report.

In light of the importance of the HTDS and future CDC reports to the public on radiation, the subcommittee suggests that the agency hold a workshop of selected risk-communication experts, scientists, journalists, and members of citizen groups to

discuss some of the important communication issues that have been raised in this case and the complex topic of the advisability of releasing unreviewed draft reports to the public. In particular, such a workshop would help to focus the growing body of social-science risk-communication research on questions about audience response to such reports as the HTDS report and simultaneously produce new research questions for systematic study. Such a workshop could also investigate how the government relates to and discusses with the public the levels of uncertainties involved in various scientific studies, as well as alternative ways of addressing public concerns about issues like Hanford. Much still needs to be known about how members of the public use and respond to risk-related messages, including the complicated roles of trust and credibility in how audiences accept and process risk messages.

9

Summary of Responses to CDC's Questions

CDC asked the National Research Council subcommittee the following six questions regarding the HTDS, the Draft Final Report and related materials and events:

- Has the analysis been carried out appropriately and completely?
- Are the presentation and the discussion of results complete?
- Are the conclusions reasonable?
- Was the material accurate and appropriate in providing guidance to the public in understanding the study findings?
- If these messages about findings need to be amended, how should the revised messages best be communicated to the public?
- With regard to release of future study reports, how can CDC improve the public communication process?

This chapter presents the subcommittee's answers to those specific questions.

HAS THE ANALYSIS BEEN CARRIED OUT APPROPRIATELY AND COMPLETELY?

The quality of the data used in the analysis is as important as the analysis itself. Our overall assessment is that the epidemiologic design was of excellent quality. The sample was based on an almost complete census of eligible subjects born in the selected years and geographic regions. Efforts to solicit study participation were thorough and appeared comparable for those with high and low doses; as a result, they achieved similar participation rates across the dose range. The HTDS was successful in obtaining subjects willing to participate, and it essentially met its goals with regard to sample size and thyroid-dose distribution.

Care was taken to maintain blinding wherever possible in the study to minimize the potential for selection bias, interviewer-induced response bias, and clinical-examination bias. The clinical examinations and laboratory studies were performed with modern methods of detecting and defining thyroid disease, although more quality-assurance procedures with regard to the cytopathology data would have been desirable. In short, the epidemiology and clinical parts of the HTDS were designed and conducted with great care. The study appears to compare extremely well with other epidemiologic studies in those respects.

The subcommittee believes that the methodology used by the HEDR project to estimate thyroid doses and their uncertainties is structurally sound. The HEDR models have been subjected to numerous reviews, and the various codes have been tested by the HEDR project staff independently of the developers to ensure correct implementation. The dose assessment has been found, on the whole, to be reasonably sound for the estimation of thyroid doses for representative, hypothetical individuals. However, errors have been found, and doubts have been raised about the validity of the results for particular environmental conditions and for the estimation of thyroid doses for specific individuals.

Although there are some fairly minor errors in the model calculations (use of sagebrush measurements to estimate pasture grass concentrations, use of wet rather than dry grass weights, and possible failure to consider wet deposition on pasture grass), the dose estimates for the period 1944–1947, the period of greatest exposure, appear reasonable. There might have been some underestimation of doses for later years, as pointed out by Hermann and Hermann (1996) and Hoffman and others (1999). However, those possible errors would have relatively modest impact because the later doses were much smaller than those in 1944–1947 and the study children were older in the later period; other studies have shown that the thyroid is much more sensitive to cancer induction in early childhood than in later childhood and adolescence.

There is reason to believe that the dose uncertainties that the HEDR project estimated and that the HTDS study used are underestimates of the total uncertainty. Some sources of uncertainty in the ^{131}I pathway to humans probably were not included, such as cow-feeding practices and commercial milk-distribution patterns. But, owing to the scattered nature of the information on uncertainties, as opposed to its being summarized in one source, the subcommittee could not be sure about the uncertainties. In addition, uncertainties in the residential histories, lifestyle, and, especially, milk-drinking habits of the children were not accounted for.

One major component of the determination of individual ^{131}I exposures was the milk-drinking habits of the study subjects. Although an attempt was made to interview a parent or surrogate to obtain recollections of the milk consumption of each subject in childhood, it was possible to do so for only 62% of the subjects; in the other cases, default assumptions were used in calculating thyroid dose. The defaults that the CIDER model used proved to be considerable overestimates of the average doses derived from the reported milk consumption and source in the interviews. The potential impact of the discrepancy between

reported and imputed milk consumption on the dose-response analyses needs to be evaluated further.

The basic objective of the statistical analysis was to determine whether there was a cause-effect relationship between the occurrence of various thyroid diseases and the magnitude of exposure to ^{131}I released from Hanford in the period in question. That was appropriately addressed by modeling the relationship between dose and the probability of occurrence of a thyroid disease. The HTDS also considered a reasonable set of potential confounding variables for thyroid disease.

However, there was overreliance on the maximum-likelihood fitting of the linear dose-response model; for several of the important outcome variables (such as thyroid carcinoma), the model failed to converge. When the linear model as described failed to converge, an analysis could have been conducted with four or five dose groups and the average value of dose in each category could have been used as the predictor variable. That would probably have resulted in successful convergence of the model and retained reasonable power to detect an effect.

Dose-response analyses with stratification on geostrata are needed because the HTDS tabulations showed that rates of several types of thyroid disease tended to be higher in areas with low fallout. That means that the geostratum differences would induce a negative association between ^{131}I and thyroid-disease rates and might have masked a positive association between thyroid dose and disease.

The HTDS investigators took care to examine the results for study participants who proved never to have been in the dosimetry area during the time of ^{131}I exposure (the out-of-area participants). They performed sensitivity analyses to determine the impact of possible dose misestimation for those subjects and found it to be small. However, the HTDS investigators made no attempt to model the out-of-area doses for persons who were included in the main analyses. That is, if a person was in the dose-assessment area for only part of the time when there were ^{131}I releases, they

calculated his or her dose only for the time that he or she was in the dosimetry area. They implicitly assumed that the dose was zero for any time when the person did not reside in the area. No attempt was made to conduct a sensitivity analysis to evaluate how that assumption could have affected the results. Their approach to this issue could have led to attenuated or biased results in that it potentially estimated total fallout doses for some people but only partial doses for others.

The HTDS investigators performed an adjusted comparison of the number of thyroid cancers found in the study versus the number that would be expected in the general US population (without radiation exposure) and found no difference. But there was a great deal of uncertainty in the comparison because the degree to which thyroid screening alters the number of thyroid cancers found is not well known. The subcommittee believes that the HTDS emphasis on analyses of subjects in the study rather than on comparisons with the general population is appropriate.

There are sizable uncertainties in the doses reconstructed for individuals because of residential and, especially, dietary histories and because of the imprecision related to the source term, meteorologic conditions, pasture deposition, milk ^{131}I, source of milk, and iodine metabolism needs to be taken into account. It seems clear that analyses need to address the uncertainties explicitly, and the confidence intervals and the strength of the conclusions have to reflect them.

ARE THE PRESENTATION AND THE DISCUSSION OF RESULTS COMPLETE?

One of the difficulties that the subcommittee encountered while reviewing the work of the HTDS is that the method that was used to calculate the doses is not clearly described in the documents that it was given. That information presumably is scattered in the large number of documents that were prepared by the HEDR project and the HTDS in the course of their work. It

would be helpful to have access to a single document in which the HEDR work is summarized, the results provided to the HTDS are clearly described (with their strengths and weaknesses), and the ways in which the HTDS made use of the HEDR results to estimate individual thyroid doses are described in detail.

It is unclear to the subcommittee how the uncertainties in the thyroid doses were estimated by the HTDS and the degree to which they might have been underestimated. In part, the lack of clarity reflects the lack of a single source that describes the dose modeling, including the coefficients and the uncertainty factors. There is a need to address explicitly the magnitude of the uncertainties associated with residential and dietary histories. There is a substantial description of how the dosimetry-related data were collected from people but relatively little information on how they were used. The input into the CIDER program is described as "scenarios", but these are not explicitly described, in particular how the scenarios were constructed from the data. There are a number of references to the use of default values in the CIDER program, but there is no discussion of which parameters used default values or of the degree to which default values changed as life circumstances changed for a given person (for example, if a person moved from a farm to a city).

A tabular presentation of the pathways to diagnosis would help the reader to assess how the final diagnoses were assigned. Assignments were made in more than one way for each of the clinical outcomes. A table for each diagnosis with a list of the methods of diagnosis and the number of times each was used should be included in the full report, and these data should be looked at for indications of unsuspected ascertainment bias.

Dosimetry-error issues apparently were not fully treated in the analysis of the study power. The same issue is raised by the results; in particular, ignoring dosimetry error could lead to unrealistically narrow estimates of the confidence limits that should be applied to the estimated parameter values. It is unlikely that the estimated dose-response relationships would change in an

important way, but confidence intervals that take dosimetry error into account would provide further information about the uncertainty of risk estimates.

Some results were presented in an abstract, rather uninformative manner. For example, there was a scatterplot of individual thyroid doses on a logarithmic scale, but no table of the frequency distribution of doses, which would have meant more to readers. Similarly, one expects to see in radiation-epidemiology reports tables that show observed and expected numbers of disease outcomes in, say, four or five dose groups. Such key tables were notably absent from the HTDS Draft Final Report.

A description of the estimated dose distribution (distribution of median doses for the individuals in the study) according to such important categories as geostratum, year of birth, and amount of milk consumption in childhood would be helpful, especially in interpreting the finding that the least exposed geostratum appeared to have the highest rates of many of the thyroid diseases or abnormalities.

It was not very clear from the report how confounders of the dose-response relationship were treated, and results adjusted for possible confounders were rarely given. The HTDS investigators conducted analyses of the various thyroid-disease end points to evaluate a number of possible risk factors for confounding effects or effect modification, but they presented no tables to show a summary of the results of these analyses for any of the disease end points. Of particular value would be a presentation of results stratified by sex, age at initial exposure, magnitude of NTS and global fallout, and history of substantial medical radiation exposure. A tabulation of the number of study participants who were out of the dosimetry area for part of the exposure period is also needed.

It has been suggested that the investigators should have given more attention to comparing the rates of thyroid disease that they found with the rates in other, unexposed populations. However, comparisons with an external, general population are

fraught with problems in that people living in various geographic areas vary in their risk of thyroid diseases because of dietary and other factors and the rates of detected disease depend heavily on the frequency and sensitivity of thyroid examinations in the population. Those circumstances produce screening effects that cannot be readily disentangled so as to permit meaningful comparisons with rates from other geographic regions that did not have comparable screening. Therefore, the subcommittee believes that the HTDS investigators rightly chose to emphasize the internal comparisons rather than general population comparisons.

The discussion of the results was seriously incomplete in that it said little about whether the confidence intervals were wide enough to be compatible with those of other, parallel studies. Most important, there was no adequate discussion of how dosimetric uncertainties might have affected the confidence intervals and the statistical power of the study.

In the statistical-power analyses, it was assumed that the dose uncertainties were all of the "Berkson type"—a type of measurement error that does not affect statistical power. However, if a substantial, fraction of the variability of the HEDR individual dose estimates actually is due to non-Berkson error or to multiplicative errors, or if there is a substantial additional component of error due to uncertainties in milk consumption, lifestyle, and residential history, the power of the study might have been reduced below a point that would normally be considered acceptable. Furthermore, the apparent heterogeneity among geostrata might also have reduced statistical power. In contrast, if the doses were systematically underestimated, as has been alleged by others, that might tend to increase the statistical power of the study.

ARE THE CONCLUSIONS REASONABLE?

Our overall assessment is that the design and execution of the epidemiologic-clinical study was appropriate to the task. The sample was based on an almost complete census of eligible

subjects born in selected years in what are believed to be the high-dose regions and inclusion of subjects born in the same years and living in regions where the doses from Hanford releases were lower. Efforts to elicit study participation were thorough and appeared comparable for those with high and low doses. Care was taken to avoid methodologic study biases. The clinical examinations and laboratory studies used the best modern methods of detecting and defining thyroid disease. Overall, the authors deserve high marks for the carefully conducted and documented epidemiologic and clinical work they performed.

To the degree that the dose is underestimated, the imputed risk estimates will be too large; but systematic, across-the-board dose underestimation does not alter the statistical significance of dose-response trends. Hence, in this study—in which, as it turned out, the primary issue became whether there is any association between ^{131}I exposure and thyroid diseases—the impact of possible dose underestimation might not change the study conclusions appreciably, except for two caveats. First, if there was an across-the-board underestimation of doses, the true statistical power of the study would have been greater than one would estimate it to be, given the reported dose distribution; the negative results would be more persuasive than they are. Second, if the doses were underestimated more for some study subjects than for others, this would, in effect, act as another source of measurement error that would tend to cancel out the gain in statistical power achieved by having generally higher doses. Therefore, a simple generalization about the effects of dose underestimation cannot be given.

If the doses were systematically overestimated, the statistical power of the study would be less than claimed.

In both versions of the analysis of mortality data—whether data are arrayed by year of birth or year of death—SMRs for perinatal conditions and congenital anomalies in the period before the Hanford releases (that is, before 1945) are increased about as much as the SMRs in the period when exposure

occurred. That constitutes evidence that exposure to ^{131}I is probably not responsible for the increased mortality from thyroid disease in the Hanford study group, although some further analysis is recommended.

The confidence intervals for the risk of thyroid cancer in this study and the Utah NTS study overlap to some degree. Moreover, on the basis of the considerations above, it is evident that the confidence intervals for the HTDS in fact depend on dosimetry-error assumptions; measurement errors in the dosimetry of the type that would attenuate the dose-response association, which we believe to be present, would cause the confidence intervals for the HTDS to be appreciably wider than the ones based on the information in the Draft Final Report.

Uncertainties in the magnitude of individual thyroid dose estimates and the relatively small sample in this study limit the generality of the conclusions that one can draw from the study regarding the magnitude of thyroid-cancer risk to other populations exposed to similar doses from ^{131}I. Data on thyroid-cancer risk are awaited from the large number of thyroid cancers observed in children exposed to high doses from the Chernobyl accident.

The statistical power of the HTDS to detect dose-response relationships might have been overestimated because effects of the types of dosimetry error that would attenuate the associations were ignored. The evidence of heterogeneity of many of the thyroid diseases or abnormalities among geostrata also suggests that the power of the study was weakened by geographic variations in unmeasured or unknown confounders that affected the outcomes. Such reductions of statistical power lower our appraisal of the study in relation to other studies and of the use of this study's results for making predictions about disease risk in other populations exposed to ^{131}I at low doses and at low dose rates.

The results of the HTDS are fundamentally important to the population living around Hanford. The study examined a substantial fraction of the highly exposed population and failed to find direct evidence of an effect of Hanford exposure on thyroid

disease risk; nor was there evidence of abnormally high rates of thyroid disease in the Hanford "downwinder" sample as a whole. It is very likely that no other reasonable study of this population could have found such effects—not only for technical reasons (such as difficulties in reconstructing doses), but because factors other than the Hanford releases appear to be more important in determining the amount of thyroid disease in the population. The fact that the two counties believed to have the least exposure tended to show higher rates of thyroid disease than the most exposed counties bears witness to that. Considerations of basic factors, such as the prevailing wind direction and distance from the Hanford site, indicate that those two counties should have had less ^{131}I exposure than the other counties in the study, but their disease rates were higher. That implies that there was not a strong association between ^{131}I exposure and thyroid disease. Nevertheless, neither this evidence nor the negative dose-response results rule out the possibility of a weak association.

We have already noted that the HTDS design and conduct were as good as they could have been, given the size and distribution of the population at risk and the long time between exposure and the study. The thorough examination of the highest-risk group—the youngest children living in the highest-dose regions—failed to find an increased incidence of thyroid cancer, the hallmark effect of high thyroid doses from ^{131}I. Even if some of the factors that affected the distribution of dose (for example, milk consumption) were poorly measured, they were as well measured in this study as is possible 40–50 years after the fact.

In evaluating the HTDS, it is useful to distinguish what the subcommittee regards as two aims of the study. The first was the determination of whether patterns of thyroid morbidity among those in the study region during the fallout period correspond to likely patterns of exposure in the HTDS study sample irrespective of specific estimated doses, the likely patterns of exposure being based principally on location of residence during early childhood (distance down wind) and milk-drinking habits. This study, by

virtue of its size and how well contact and screening were completed, appears to have had the ability to address that aim effectively, although the presentation of results could be improved. The second aim was to use assigned HEDR doses to estimate dose-response relationships. Many of the caveats discussed in this report about the effect of uncertainties in doses on the power of the study, confidence intervals for the estimated dose-response relationships, and the correspondence of the HTDS results with those of other studies are related to the second aim. The HTDS might have substantially overestimated its ability to assess a dose-response relationship, because of unallowed-for uncertainties, both systematic and random, in the HEDR doses.

The absence of a thyroid-cancer risk above expectation, is based on a comparison of rates between high-and low-dose regions and a dose-response analysis that used dose-reconstruction methods and found no increased risk but did have wide uncertainties in the reconstructed-dose estimates and hence in the estimated risk coefficients. From those observations, one can only state that at face value the HTDS was negative. If an exposure of a population to ^{131}I radiation of magnitude similar to that estimated in the HTDS were to occur elsewhere, one cannot predict with confidence whether an increase in risk would be seen. The small number of thyroid-cancer cases and the wide uncertainties in individual doses afford little confidence in the risk estimates derivable from this study.

At the time of the initial release of the Draft Final Report, the HTDS investigators indicated that residents of downwind areas should feel relief that their proximity to the Hanford nuclear site did not result in increased risk of any thyroid morbidity. Such statements are, according to the arguments above, reasonable in specific instances. For example, a healthy 55-year-old former resident of the area near the Hanford site (say, Benton County) who remembers drinking a large amount of milk as a child can take comfort in learning that there is no evidence that his or her risk of thyroid morbidity is higher than that of other subjects in

the general HTDS study area. The HTDS, in fact, appears to be definitive on that point for most of the outcomes considered.

In contrast, for good reasons discussed above, the HTDS was not designed to compare thyroid morbidity rates among the entire population downwind from Hanford with those among populations living elsewhere. At the various public-comment meetings, people who lived in down-wind areas stated that they and their families experienced more thyroid disease than would have been expected in the population at large. They could be right, and their disease might have been the result of unusual fallout or ingestion patterns. However, thyroid disease tends to run in families, and the particular occurrences could be related to genetic factors in the families, chance occurrences, or even mistaken diagnoses. The small number of thyroid cancers detected in the HTDS was in line with the background rates initially estimated for the study, but the presence of screening effects makes it impossible to compare this intensively screened sample with any other similar population. Other thyroid conditions were much more commonly detected in the sample, but again no comparison group has undergone similar degrees of thyroid screening. The lack of evidence of a dose-response relationship for any of these conditions suggests—but does not prove—that the overall risks were unaffected by Hanford releases. The evidence does not rule out (although it provides no support for) a weak association that could affect, for example, those already susceptible to thyroid disease because of predisposing genetic factors.

Thyroid cancer is not a common disease, and it would be reasonable in future epidemiology surveys to identify, document, and investigate clusters by using molecular-biology probes to characterize genetic polymorphisms that could make people more sensitive to ionizing radiation or to look for oncogene prevalence in affected groups. These methods are developing rapidly, and it is likely that they will play a role in future environmental-epidemiology studies.

WAS THE MATERIAL ACCURATE AND APPROPRIATE IN PROVIDING GUIDANCE TO THE PUBLIC IN UNDERSTANDING THE STUDY FINDINGS?

For the most part, the written messages in the Draft Final Report, those about the report in public documents, and the messages given orally by the HTDS investigators at the media and public briefings were accurate, but they were sometimes inappropriate and misleading because they included statements that were too positive, definite, and strongly worded, given the uncertainties that applied to the study. In addition, the uncertainties related to the study were not discussed and should have been. Clearly, those problems contributed to the public upset that resulted from the release of the Draft Final Report.

However, the problems related to the report's release should not overshadow the attempts made over the years to inform and involve members of the public about the study. They were valuable and generally done well. Such information channels as newsletters, background fact sheets, and a Web site were good ways of trying to reach members of the public directly with information about the study, and they augmented the information in the mass media. The subcommittee recommends that similar public communication efforts be continued regarding this report and others by CDC. Keeping the study process and activities as "transparent" as possible for the public is valuable and should not be abandoned because of the problems encountered with the release of the Draft Final Report of this study.

IF THESE MESSAGES NEED TO BE AMENDED, HOW SHOULD THE REVISED MESSAGES BEST BE COMMUNICATED TO THE PUBLIC?

Some of the overly positive messages given out at the January briefing have already been softened through CDC efforts during public meetings in Spokane and Seattle in May 1999. Handouts, a summary of the findings, and other written materials prepared by CDC also have conveyed similar information and made more of an effort to explain that no epidemiologic study can

determine "whether an individual person's thyroid disease is or is not caused by Hanford radiation exposure" (Cary, 1999).

Despite substantial attempts to publicize the May 1999 meetings, including paid advertising, they were not well attended by the public. So the "softer" CDC message was not delivered widely, although at least one newspaper did run a story about the new CDC approach. Several factors can account for the low attendance at the two meetings, including the period that had elapsed between the January release of the report and early May, the perception that no new information would be gained at these meetings, and the greater distance of the two meeting sites from Hanford, compared to Richland, the site of the January briefing. Another interpretation might be that both the HTDS and CDC have lost some credibility on this issue and that people chose not to listen to what they had to say. All those factors and others might have been operating, but the last one should have serious consideration and study before release of the final report. If credibility has been lost, plans must be developed to try to restore it to some degree, particularly for the release of the final report.

Given all the communication problems that resulted from the release of the Draft Final Report, the subcommittee recommends that another detailed communication plan be developed for release of the final report. To the greatest extent possible, those working on the plan should brainstorm about unexpected situations like those which affected the release of the draft report and devise plans to handle them more effectively. It is imperative that messages from the final report take into account the various audiences being addressed and show concern and sensitivity for the thyroid-health issues that people perceive as affecting them. Any changes made to the Draft Final Report must be clearly outlined and explained, including why they were made, what group suggested that they be made, and what impact they have had on the results of the HTDS. Every reasonable effort must be made to present the full picture of the study results, including all the uncertainties and other problems.

CDC and the HTDS personnel should work together on the wording of the messages for the public about the final findings of study. Academic freedom is of great concern in letting researchers give their interpretation of the data by other groups, such as CDC, should not be offered. It would be better to reach a consensus; if it cannot be done, showing differing interpretations is preferable to having one side dominate the other to the point that a government agency overrides the predominant point of view of the principal investigators, or vice versa. Such advice goes against some risk-communication dogma that states that the message should be kept simple and that all parties involved should show a united front. Sometimes that works, but often when members of the public are actively engaged in and educated about an issue, such a simple approach backfires, as it did with the release of the Draft Final Report.

A plan to brief the active citizen groups should be developed so that they have enough information to be able to respond to media inquiries about the report. A much simpler briefing structure should be devised to make the information available quickly to as many people as possible. In addition, an embargoed release of the report to journalists should be arranged so that they have a few days to read through it and develop informed questions before the briefing. Many government agencies and scientific organizations do that routinely and get much better, informed coverage because of it.

As for details, the executive summary of the final report should be edited carefully to eliminate unnecessary technical jargon and complexity, which made it difficult for even educated readers to understand portions of the draft Final Report. All written press releases, handouts, and other materials should be dated and numbered for easy reference; that was not always done with the material for the draft report. And all briefings should be videotaped and transcribed for future reference by CDC, the HTDC investigators, and others.

Little information has been available about the quality of communication to people who participated in the HTDS. A random sample of those people should be surveyed about their responses to the communication efforts carried out in connection with the draft report to seek ways to improve communication with them about the final report. The study participants seem to be an overlooked group; they need to be brought more into the communication loop.

WITH REGARD TO RELEASE OF FUTURE STUDY REPORTS, HOW CAN CDC IMPROVE THE PUBLIC-COMMUNICATION PROCESS?

Attempts to establish and maintain an information blackout before release of the Draft Final Report ran counter to the previous spirit of information-sharing with citizen groups in the region. The blackout, although somewhat understandable and probably policy-driven, led to ill feelings among citizens who had worked closely with the HTDS and CDC, and it led to a leaked report. Trying to maintain such a blackout during *multiple* briefings—particularly in Washington, DC—is something that CDC should reconsider when a controversial report of great public interest is involved. If nothing else, the subcommittee recommends that the briefing structure be simplified and that citizens who have participated in the advisory process all along be given higher priority in the briefing structure.

Multiple conference-call briefings, which are relatively inexpensive, appear to be ineffective. None of those who participated in them and provided information for the present report liked them. They have many disadvantages, and the subcommittee suggests that different ways be devised to brief groups at different locations. Briefings by satellite, although expensive, might pay off in the long run in more effective communication. Other possible uses of advanced communication technology, such as the World Wide Web, also should be considered. A lower-technology way to brief groups in different areas is to have simultaneous briefings led by different people

involved in the study. That method has its problems, but they seem less serious than those posed by conference-call briefings.

Many of the suggestions made for handling the final HTDS report can also be applied to future CDC reports and need not be restated here. However, the subcommittee feels that it would be helpful for the CDC to ask a group of about a dozen people to come together during a 1- or 2-day workshop to discuss how the agency can improve its public-communication process and the release of future reports.

Some of the people invited to such a workshop should be experts in risk communication, but there also should be a wider representation, including journalists, outside scientists who have worked on CDC studies, and members of citizen groups who have served on CDC advisory boards. Given careful planning, perhaps with a few case studies distributed to participants in advance, such a workshop might help CDC to avoid some of the pitfalls that occurred with the release of the HTDS Draft Final Report.

One point for such a group to consider is the advisability of publicly releasing draft reports before external peer review and, if such a release is required by law or contract, how to do so effectively. Although circumstances in this particular situation might have forced the HTDS investigators and CDC to release a draft report to the public and the media, it is problematic to have preliminary information conveyed to the public if external peer review could suggest important changes in the final report. If substantial changes are made, what will the public think about the study and the investigators? That their initial results were not scientifically valid? That they raced to release information that was politically favorable to the government and not to citizens in the region? This is not a criticism of the actions that occurred, but a call to think about these issues more carefully, not only for this project, but also for others. Release of draft scientific studies, although it potentially serves the information needs of the public, also has the potential to cause confusion and to undermine the credibility of researchers and government agencies. Of course,

keeping information from the public can do the same thing, so this is not an easy issue to deal with. It should, however, be carefully considered not only by CDC, but also by other government agencies, Congress, the National Academy of Sciences and National Research Council, and the scientific community as a whole. Science journalists, too, might need to re-evaluate how to apply their own guidelines that advocate writing primarily from peer-reviewed studies and reports as they are faced with more and more cases in which government agencies and others announce the results of unreviewed draft reports at news conferences.

10

REFERENCES

Anderson, D.M., Bates, D.J., Marsh, T.L. (1993). Estimation of 1945 to 1957 food consumption. Pacific Northwest National Laboratory: Richland, WA.

Associated Press. (1999). Study disputes Hanford poisoned people. In *The Salt Lake Tribune* pp. A12: Salt Lake City, UT.

Astakhova, L.N., Anspaugh, L.R., Beebe, G., Bouville, A., Drozdovitch, V., Garber, V., Gavrilin, Y., Khrouch, V., Kuvshinnikov, A., Kuzmenkov, Y., Minenko, V., Moschik, K., Nalivko, A., Robbins, J., Shemiakina, E., Shinkarev, S., Tochitskaya, S. and Waclawiw, M. (1998). Chernobyl-related thyroid cancer in children of Belarus: a case-control study. *Radiat. Res*., 150, 349–356.

Beck, D.M., Darwin, R.F., Erickson, A.R., Eckert, R.L. (1992). Milk cow feed intake and milk production and distribution estimates for Phase I. Pacific Northwest National Laboratory: Richland, WA.

Budowsky, J., Piro, J.D., Zegarelli, E.V., Kutscher, A.H., Barnett, A. (1956). Radiation exposure to the head and abdomen during oral roentgenography. J. Am. Dental Ass. 52, 555–559.

Carroll, R.J., Ruppert, D., Stefanski, L.A. (1995). *Measurement Error in Nonlinear Models*. Chapman, Hall: London.

Cary, A. (1999). Telephone interview.

CDC. (1999a). Congressional Briefing Document on Hanford Thyroid Disease Study, "Summary of the Study and the Primary Findings".

CDC. (1999b). Summary of the Preliminary Results "The Hanford Thyroid Disease Study Draft Final Report". Centers for Disease Control and Prevention: Atlanta, GA.

Davis, S. (1999a). Quoted in CDC Press Release, "Draft Report: Results from the Hanford Thyroid Disease Study".

Davis, S. (1999b). Telephone interview.

Davis, S., Kopecky, K. (1998). Executive Summary, HTDS Draft Final Report. Fred Hutchinson Cancer Research Center: Seattle, WA.

Davis, S., Kopecky, K.J., Hamilton, T.E. (1999). Letter to the editor. In *Seattle Post-Intelligencer*: Seattle, WA.

Day, N.E., Byar, D., Green, S. (1980). Overadjustment in case-control studies. *Am. J. Epidemiol*., 112, 696–706.

Deonigi, D.E., Anderson, D.M., Wilfert, G.L. (1994). Commercial milk distribution profiles and production locations. Pacific Northwest National Laboratory: Richland, WA.

Dunning, D.E., Schwarz, G. (1981). Variability of human thyroid characteristics and estimates of dose from ingested 131I. *Health Phys*., 40, 661–675.

Dwyer, T.J., Gardner, J., Halvorsen, K., Krall, E.A., Cohen, A., Valadian. (1989). Memory of food intake in the distant past. *Am. J. Epidemiol*., 130, 1033–1046.

Farris, W.T., Napier, B.A., Eslinger, P.W., Ikenberry, T.A., Shipler, D.B., Simpson, J.C. (1994). Atmospheric pathway dosimetry report, 1944–1992. Pacific Northwest National Laboratory: Richland, WA.

Farris, W.T., Napier, B.A., Ikenberry, T.A., Shipler, D.B. (1996). Radiation doses from Hanford site releases to the atmosphere and the Columbia River. *Health Phys*., 71, 588–601.

FHCRC. (1999a). Hanford Thyroid Disease Draft Final Report. Fred Hutchinson Cancer Research Center: Seattle, WA.

FHCRC. (1999b). Summary Final Report of the Hanford Thyroid Disease Study. Fred Hutchinson Cancer Research Center (Seattle, WA) and Centers for Disease Control and Prevention (Atlanta, GA).

Goble, R.L. (1968). Comparison of a calibration of the Hanford Environmental Dose Reconstruction model suite to the validation data sets and to the model with its initial parameter choice. Center for Technology Environment and Development, Clark University: Worcester, MA.

Goodman, M.T., Kolonel, L., Wilkens, L. (1992). The association of body size, reproductive factors and thyroid cancer. *Br. J. Cancer*, 66, 1180–1184.

Hamilton, P. and others (1989). A follow-up study of persons who had iodine-131 and other diagnostic procedures during childhood and adolescence. Rockville, MD: FDA Technical Report.

Hanford Health Effects Subcommittee. Meeting transcripts (excerpts). (1998). Salt Lake City, UT.

Heeb, C.M. (1993). Iodine-131 Releases from the Hanford Site, 1944 through 1947. Pacific Northwest National Laboratory: Richland, WA.

Heeb, C.M. (1994). Radionuclide Releases to the Atmosphere from Hanford Operations, 1944–1972. Pacific Northwest National Laboratory: Richland, WA.

Hermann, B., Hermann, F.J. (1996). Iodine-131 Release to the Atmosphere from Hanford Separations Plants for the Period May 1948 through December 1960. (A report prepared on behalf of plaintiffs in Hanford litigation.) 10 March, 1996.

Hoffman, F.O., Amaral, E., Mohrbacher, D.A., Deming, E.L. (1988). The comparison of generic model predictions with Chernobyl fallout data on the transfer of radioiodine over the air-pasture-cow milk pathway. *J. Environ. Radioactivity*, 8, 53–71.

Hoffman, F.O., Nair, S.K., Wichner, R. (1999). Review of HEDR Source Term for the Atmospheric Releases of I-131 .

SENES Oak Ridge, Inc., Center for Risk Analysis: Oak Ridge, TN.

Hoffman, F.O., Thiessen, K.M., Frank, M.L., Blaylock, B.G. (1992). A quantification of the interception and initial retention of radioactive contaminants deposited on grass by simulated rain. *Atmos. Environ*., 26A, 3313–3321.

Holm, L. (1991). Cancer risks after diagnostic doses of ^{131}I with special reference to thyroid cancer. *Cancer Detection and Prevention* , 15:27–30.

HTDS Newsletter. (1999). Questions and Answers about the HTDS Results. Fred Hutchinson Cancer Research Center: Seattle, WA

ICRP. (1990). Age-dependent doses to members of the public from intake of radionuclides: Part 1. Pergamon Press: New York.

Jacob, P., Goulko, G., Heidenreich, W.F. (1998). Thyroid cancer risk to children calculated. *Nature*, 392, 31–32.

Kerber, R.A., Till, J., Simon, S., Lyon, J., Thomas, D., Preston-Martin, S., Rallison, M., Lloyd, R., Stevens, W. (1993). A cohort study of thyroid disease in relation to fallout from nuclear weapons testing. *J. Am. Med. Assoc*., 270, 2076–2082.

Kolonel, L.N. (1985). Cancer incidence among Filipinos in Hawaii and the Philippines. *Monogr. Nat. Cancer Inst*., 69, 93–98.

Krimsky, S., Plough, A. (1988). *Environmental Hazards*. Auburn House Publishing: Dover, MA.

Lee, W., Chiacchierini, R., Shleien, B., Telles, N. (1982). Thyroid tumors following I-131 or localized X irradiation to the thyroid and pituitary glands in rats. *Radiat. Res*., 92, 307–319.

McTiernan, A., Weiss, N., Daling, J. (1987). Incidence of thyroid cancer in women in relation to known or suspected risk factors for breast cancer. *Cancer Res*., 47, 292–295.

Nagataki, S., Shibata, Y., Inoue, S., Yokoyama, N., Izumi, M., Shimaoka, K. (1994). Thyroid diseases among atomic bomb survivors in Nagasaki. *J. Am. Med. Assoc*., 272, 364–370.

Napier, B.A., Simpson, J.C., Eslinger, P.W., Ramsdell, J.V., Thiede, M.E., Walters, W.H. (1994). Validation of HEDR Models. Pacific Northwest National Laboratory: Richland, WA.

NAS/IOM. (1999). *Exposure of the American People to Iodine-131 from Nevada Nuclear-bomb Tests: Review of the National Cancer Institute Report and Public Health Implications*. National Academy Press: Washington, DC.

National Research Council. (1989). *Improving Risk Communication*. National Academy Press: Washington, DC.

National Research Council. (1998). *Letter from Chairman Jack Schull to James Smith, 27 July 1999*. Board on Radiation Effects Research, National Academy of Sciences, National Research Council: Washington, DC.

National Research Council. (1999). *Letter from Chairman Jack Schull to James Smith, 25 January 1999*. Board on Radiation Effects Research, National Academy of Sciences, National Research Council: Washington, DC.

NCI. (1997). Estimated exposures and thyroid doses received by the American people from iodine-131 in fallout following Nevada atmospheric nuclear bomb tests. Volume 1. National Cancer Institute, US National Institutes of Health: Bethesda, MD.

NCRP. (1985). *Induction of Thyroid Cancer by Ionizing Radiation, Report No. 80*. National Council on Radiation Protection and Measurements: Bethesda, MD.

Nolan, W.E. (1953). Radiation hazards to the patient from oral roentgenography. *J. Am. Dental Ass*., 47, 681–684.

Preston-Martin, S., Jin, F., Duda, M., Mack, W. (1993). A case-control study of thyroid cancer in women under age 55 in Shanghai (People's Republic of China). *Cancer Caus. Cont*., 4, 431–440.

Price, D.W. (1994). Dairy Cow Feeding Regimes for Representative Doses: HEDR Project. Pacific Northwest National Laboratory: Richland, WA.

Ramsdell Jr., J.V., Simonen, C.A., Burk, K.W. (1994). Regional Atmospheric Transport Code for Hanford Emission Tracking (RATCHET). Pacific Northwest National Laboratory: Richland, WA.

Ron, E., Lubin, J., Schneider, A. (1992). Thyroid cancer incidence (letter). *Nature*, 360, 113.

Ron, E., Lubin, J., Shore, R., Mabuchi, K., Modan, B., Pottern, L., Schneider, A., Tucker, M., Boice, J. (1995). Thyroid cancer after exposure to external radiation: a pooled analysis of seven studies. *Radiat. Res*., 141, 259–277.

Ron, E., Modan, B., Preston, D., Alfandary, E., Stovall, M., Boice, J. (1989). Thyroid neoplasia following low-dose radiation in childhood. *Radiat. Res*., 120, 516–531.

Salvini, S., Hunter, D.J., Sampson, L., Stampfer, M.J., Colditz, Rosner, Willett. (1989). Food-based validation of a dietary questionnaire: the effects of week-to-week variation in food consumption. *Int. J. Epidemiol*., 18, 858–867.

Sever, L.E., Hessol, N.A., Gilbert, E.S., McIntyre, J.M. (1988). The prevalence at birth of congenital malformations in communities near the Hanford site. *Am. J. Epidemiol*., 127, 243–254.

Shipler, D.B., Napier, B.A., Farris, W.T., Freshley, M.D. (1996). Hanford Environmental Dose Reconstruction Project-- an overview. *Health Phys*., 71, 532–544.

Shore, R. and others, (1993). Thyroid cancer among persons given x-ray treatment in infancy for an enlarged thymus gland. *American Journal of Epidemiology*, 137:1068–1080.

Simon, S.L., Lloyd, R., Till, J., Hawthorne, H., Gren, D., Rallison, M., Stevens, W. (1990). Development of a method to estimate thyroid dose from fallout radioiodine in a cohort study. *Health Phys*., 59, 669–691.

Snyder, S.F., Farris, W.T., Napier, T.A., Ikenberry, T.A., Gilbert, R.O. (1994). Parameters used in the environmental Pathways and radiological dose modules (DESCARTES, CIDER, and CRD codes) of the Hanford Environmental

Dose Reconstruction Integrated Codes (HEDRIC). Pacific Northwest National Laboratory: Richland, WA.

Steele, K.D. (1999). Telephone interview.

Swanson, S., Friedman, S.M., Bridgen, P., Covello, V., Slovic, P., Cohn, J. (June 1991). *Risk Communication Manual for Electric Utilities.* Electric Power Research Institute: Palo Alto, CA.

Thompson, D., Mabuchi, K., Ron, E., Soda, M., Tokunaga, M., Ochikubo, S., Sugimoto, S., Ikeda, T., Terasaki, M., Izumi, S., Preston, D. (1994). Cancer incidence in atomic bomb survivors. Part II: Solid tumors, 1958–1987. *Radiat. Res*., 137, S17–S67.

Wald, M.L. (1999). No radiation effect found at Northwest nuclear site. In *New York Times* p. A20: New York NY.

Warren, J.H. (1961). Control of I-131 Releases to the Atmosphere. Hanford Atomic Products Operations: Richland, WA.

Willett, W. (1990). *Nutritional Epidemiology*. Oxford University Press: New York.

Wong, F.L., Ron, E., Gierlowski, T., Schneider, A. (1996). Benign thyroid tumors: general risk factors and their effects on radiation risk estimation. *Am. J. Epidemiol*., 144, 728–733.

Appendix A

Subcommittee Activities

To undertake this study, the National Research Council established a subcommittee of its standing Committee on An Assessment of Centers for Disease Control and Prevention Radiation Studies to review the conduct and report of the Hanford Thyroid Disease Study (HTDS). The Subcommittee to Review the Hanford Thyroid Disease Study Final Results and Report included 11 members on its final roster.

In its request to the National Research Council, the Centers for Disease Control and Prevention (CDC) asked that the National Research Council's work be conducted in public to the fullest extent possible. Consistent with the policies of the National Academy of Sciences, the subcommittee conducted its fact-finding activities in public meetings and met in closed session only to consider findings and recommendations. At its first meeting, the subcommittee examined its composition to make certain that necessary expertise and perspectives were represented and that no important conflicts of interest or bias existed. The expertise sought in this group included thyroid disease, epidemiology, biostatistics, risk assessment, dose reconstruction, risk communication, radiation oncology, clinical practice, and public health. After a brief discussion, it was decided that two subcommittee members should not serve on the subcommittee. Lynn Anspaugh and William Schull volunteered to step down because of possible or

perceived conflicts of interest. Roy Shore was appointed to chair the subcommittee.

The subcommittee was initially charged with the three tasks outlined in the beginning sections of this report, all related to evaluating the quality of the analysis and conclusions in the draft of the HTDS findings. In April 1999, CDC, prompted by its own concerns and those of interested members of the public, asked the subcommittee to review and comment on a group of materials prepared by CDC and the Fred Hutchinson Cancer Research Center (FHCRC) and provided to the public at the time of release of the HTDS Draft Final Report or thereafter.

The subcommittee met four times in 1999: on February 4–5, in Atlanta, GA; on March 29–30, in Augusta, GA; on June 18–19, in Spokane, Washington; and on August 30–31, in Washington, DC. Two of the meetings included information-gathering sessions.

On the first day of its first meeting, in open session, the subcommittee and observers received a briefing on the findings of the HTDS Draft Final Report by Scott Davis and Kenneth Kopecky of FHCRC. Oral comments were presented by Judith Jurji and Louise Kaplan, members of the public who had lived in the Hanford area. Owen Hoffman, of Senes Oak Ridge, provided technical comments via a conference call, and written materials were distributed to those in attendance. In addition, subcommittee consultants—Maureen Hatch, director of the Division of Epidemiology and associate professor in the Department of Community and Preventive Medicine, Mount Sinai School of Medicine; and Arthur Schneider, chief of the Section of Endocrinology and Metabolism, University of Illinois College of Medicine—were in attendance to provide answers to questions related to their participation in a review of the HTDS conducted by CDC. On the second day, subcommittee members and members of the public were given the opportunity to ask additional questions on the findings of the HTDS in a half-day open session, and an afternoon session was closed for the subcommittee to develop its report and to explore its plans to respond to its charge.

The first day of the third meeting consisted of a half-day closed session for the subcommittee to discuss its preliminary views on the HTDS. The second day was a full-day session open to the public and dedicated to gathering information and the views of the public and technical experts on the implementation, the power, and the presentation of the HTDS; the Draft Final Report; and a variety of issues related to how the report was communicated to the public.

The National Research Council called on the assistance of four interested and involved citizens—Trisha Pritikin, Tim Connor, Judith Jurji, and Louise Kaplan—to assist in the planning of the second meeting and to recommend speakers to be invited. In addition, the advice of the Hanford Health Information Network (HHIN) was sought, in particular, that of Executive Director Bea Kelliegh. The HHIN later assisted the National Research Council in announcing the public meeting by mailing some 18,000 postcard notifications to interested persons in and around Idaho, Oregon, and Washington state. The invitations went to households on the HHIN mailing list and to over 100 citizen advisory and public-interest groups. About 60 people attended the meeting, including members of the press from the three states.

The public meeting was structured to solicit comment from technical experts and laypersons on four specific topics identified through consultation with the citizen advisers mentioned: thyroid dosimetxy and uncertainty, other evidence and contextual information related to Hanford exposures, statistical power and study design, and communication. Each topic was discussed for a least an hour and then in open-microphone sessions lasting for about a half-hour.

The meeting progressed through the four topics with 14 invited presentations. Attendees were encouraged to make oral statements or to provide written questions, concerns, and comments to the subcommittee. Written submissions from those who could not attend the meeting were invited.

The following were the subjects discussed and the invited speakers:

Session 1: Dosimetry and Uncertainly

Owen Hoffman (Senes, Oak Ridge)

William Farris (Pacific Northwest National Laboratory)

Trisha Pritikin

Session 2: Other Evidence and Contextual Information Related to Hanford Exposures

Keith Baverstock (World Health Organization), via teleconference

Lynn Lyon (University of Utah)

Ernest Mazzaferri (Ohio State University, Retired), via teleconference

Charles Grossman (Legacy Good Samaritan Hospital, Portland, Oregon, and Northwest Radiation Health Alliance)

Robert Schenter (Pacific Northwest National Laboratory)

Session 3: Statistical Power and Study Design

James Ruttenber (University of Colorado School of Medicine, Denver)

Tim Connor (North West Environmental Education Foundation)

James Thomas (Short, Cressman and Burgess)

Edward Liebow (Environmental Health and Social Policy Center)

Session, 4: Communication

Trisha Pritikin

Larry Jecha (Benton-Franklin Health Department)

Oral statements were presented to the committee by several people, including Darrell Fisher, Norm Buske, Linda Keir, Jude Van Buren, Juliet Van Eenwyk, Laura Chenet Leonard, Lois Camp, Gai Oglesbee, Kay Sutherland, Diane Omeron, Margaret Losh, Robert Miller, and Gordon Hilderbrand.

The meeting successfully provided an open forum whereby the subcommittee received technical statements and opinions concerning the design, implementation, analysis, and communication of the HTDS. All the information gathered became part of the National Research Council's public-access file and is available, on request, to anyone interested in it.

The third meeting of the subcommittee consisted of closed sessions to discuss the first complete draft of the subcommittee's report.

Several people were consulted at other times by subcommittee members Sharon Friedman and Susan Lederer to analyze the effectiveness of the communication of the HTDS Draft Final Report. In particular, Karen Dom-Steele, of the Spokane *Spokesman-Review*, and Annette Cary, of the *Tri-Cities Herald* provided information, as did CDC staff, including James Smith, Diana Swindle, Paul Garbe, Michael Donnelly, and Joan Morrisey.

Appendix B

Responses to Selected Comments by the Public

This section provides responses to various questions and comments posed to the subcommittee at the public meeting held in Spokane, Washington, in June 1999.

Comment: Differences in terrain were not included in the fallout model, and there were great differences in mountains and plains in that area.

Response: Major terrain effects are indirectly accounted for in the HEDR dose model even though the computer code assumed a flat terrain for the entire area being modeled. The terrain was implicitly taken into account in the meteorologic projections of where the ^{131}I plume went and to a lesser extent in a "surface roughness parameter". The hourly meteorologic data from the 16 stations in the geographic area roughly reflect the effects of the Cascade range, the Blue Mountains south of Walla Walla, and the Bitterroot Mountain range east of Spokane, Coeur d'Alene, and Lewiston. The modeling of terrain effects in a more detailed manner is impeded by the severe limitations of the meteorologic data that are available for the model for 1944–1947, the period of greatest interest.

Comment: The Spokane area has had a historically high thyroid disease incidence due to lack of iodized salt or other sources of iodine. Were the data corrected for the high incidence in the Spokane area in the 1940s and 1950s?

Response: Iodized salt was widely available by the 1940s, so low dietary iodine might have been an issue only in some unidentified and probably small subset of children in that era. Development of thyroid disease related to iodine deficiency typically takes a number of years of low iodine intake. The children in this study, born when iodized salt had already been introduced, were probably much less subject to iodine-deficiency disease than were their parents and grandparents.

However, one potential impact of low iodine in a subset of the children might have been to increase the uptake of ^{131}I, that is, to give iodine-deficient children more radioactive iodine than they otherwise would have received. In the absence of knowledge of which children were iodine-deficient, it would be impossible to factor this into the individual dose estimates.

Comment: The cohort studied should have been far larger, at least 10,000 persons. More higher doses should have been included in the cohort.

Response: The HTDS investigators performed initial calculations of statistical power that showed excellent statistical power for the set of assumptions that they used, which at the time appeared plausible; and that made them think that they did not need any more subjects. However, the NRC subcommittee believes that their assumption that dose-measurement errors were of a type ("Berkson error") that did not adversely affect statistical power is unlikely to be valid. There are several reasons to believe that the dose-measurement errors did reduce statistical power. Hence, we agree that in principle a larger study would have been very desirable.

Nevertheless, it seems unlikely that the number of high-dose study subjects could have been increased appreciably. It is preferable to study those who were young children (less than 10 years old and preferably less than 5) at the time of ^{131}I exposure, in that children, adolescents, and adults are markedly less sensitive to radiation effects on the thyroid than young children are. The HTDS investigators apparently studied all possible young children in the high-dose counties (Adams, Franklin, and Benton) unless there are other high-dose counties that the subcommittee is unaware of. The only other possibility would have been to increase the numbers somewhat in the intermediate-dose category (for example, Walla Walla County), but this would probably have increased the statistical power by only a small amount, and adding more subjects from low-dose counties would likewise have only a small effect on statistical power.

Comment: Increases in mortality and birth defects were high in the area before and after the Hanford exposure period with no explanation.

Response: Perinatal mortality (death rate during the first month of life) and mortality due to birth defects (congenital anomalies) were somewhat higher than national rates. However, as the comment noted, they were higher in the area both before and during the period of Hanford fallout exposures. Therefore, it seems not very likely that the higher rates were caused by the fallout. A more detailed study of birth defects in Hanford downwinders found no increase in birth-defect rates, except for a possible increase in neural-tube defects (Sever and others, 1988).

Comment: Risk analyses from other ^{131}I thyroid studies appear different than this study.

Response: The Utah study of thyroid disease after Nevada Test Site ^{131}I fallout showed marginal excesses of thyroid cancer, thyroid nodules, and both combined; but when dose uncertainties were properly included in the risk estimates, the results were not statistically significant. Furthermore, our subcommittee shows in its report that the thyroid-cancer risk estimates from the Utah and Hanford studies are probably statistically compatible with each other.

A comparison with the Marshall Islanders is questionable because the doses were very high for children on the islands studied and were mostly from short-lived forms of radioactive iodine and gamma rays, rather than from ^{131}I.

Two studies of ^{131}I administered to young people for diagnostic medical purposes have not shown statistically significant excesses of thyroid cancer (Holm, 1991; Hamilton and others, 1989). The average thyroid doses in the two studies were about 800 and 1500 mGy. However, most of the subjects were adolescents at the time of ^{131}I exposure, and the atomic-bomb study and other studies show that radiation exposure in adolescence causes much less thyroid cancer than the same exposure in early childhood. Therefore, it is difficult to interpret the negative results.

The Chernobyl studies in Ukraine and Belarus have shown increases in thyroid cancer after the Chernobyl ^{131}I releases. The risk per mGy is not well quantified at this point, so it is not clear whether the Chernobyl and Hanford results are statistically compatible.

Comment: The study should have investigated synergism with other environmental insults.

Response: The study did make some attempt to do so with regard to radiation, the main known environmental risk factor for thyroid disease. The HTDS investigators obtained a history of diagnostic and therapeutic medical irradiation and

information on occupational radiation exposure, and they found no synergism.

Other studies have found little evidence of synergism of radiation and other environmental exposures in causing thyroid cancer. For instance, one study that has investigated this found no synergism of oral contraceptive use, hormone-replacement therapy, and smoking (Shore and others, 1993).

Comment: A study of "clusters" should be done, particularly in families in which no previous thyroid disease had been found. Families with thyroid problems should be studied.

Response: At the various public-comment meetings, a number of people who lived in downwind areas stated their belief that they and their families had experienced more frequent thyroid diseases than would have been expected in the population at large. They could be right, and their disease could have been the result of unusual fallout or ingestion patterns. However, it is also true that thyroid disease tends to run in families, and the particular occurrences could be related to genetic factors in the families, chance, or even mistaken diagnoses. A compilation and study of such clusters could have been undertaken, but that would have been a special study and was not part of the HTDS design.

Comment: Screening effects are a major unresolved issue that needs evaluation.

Response: The HTDS investigators wanted to compare the rate of thyroid cancer among the Hanford downwinders with that found in an unirradiated general population. But to do so they knew that they needed to take account of the fact that their study population all had sensitive thyroid screening with ultrasonography and palpation of the thyroid by expert thyroid

physicians. It is well known that many more thyroid cancers and nodules are detected when there is intensive screening.

To address the inequity in thyroid-cancer or nodule detection between the intensively screened population and the general US population, the investigators chose an increase by a factor of 3 in thyroid-disease rates due to screening on the basis of an estimate in a 1985 publication (NCRP, 1985). However, two studies since then have suggested different screening factors, from 2.5 to about 7 for thyroid cancer and 17 for thyroid nodules (Thompson and others, 1994). Hence, we have much uncertainty about the size of screening effects. That is one of the reasons that it was more appropriate to compare disease rates within the study population, in which everyone underwent screening, than between this study population and some other, mostly unscreened population.

Comment: Other health problems that could possibly result from ^{131}I exposure should be included and not just thyroid disease alone.

Response: ^{131}I concentrates in the thyroid, where it remains for several weeks. The concentration and long residence time lead to potentially large doses to the thyroid. Other organs receive only about one-thousandth of the dose received by the thyroid, because they do not concentrate and retain ^{131}I or its radioactive metabolites. Except for the parathyroid glands, no other organs could have received biologically significant doses from the environmental releases from Hanford. The parathyroid glands are intimately attached to the thyroid and receive fairly high doses because of their proximity. Changes in parathyroid function were screened for, and no changes related to radiation injury were found; because no effects were seen in the parathyroid glands, it is most unlikely that radiation effects in other organs would have occurred and gone undetected.

Comment: Doses were too low to detect any thyroid changes. Only about 2 dozen in the study had estimated doses over 100 rad.

Response: Many scientists believe that the bulk of evidence suggests that even quite small doses can cause thyroid cancer. For instance, a study in Israel of children who received x-ray thyroid exposure of about 100 mGy (10 rad) had clear excesses of thyroid cancer and thyroid nodules (Ron and others, 1995). In comparison, the average thyroid dose in the HTDS was about 180 mGy (18 rad).

However, it is generally believed that ^{131}I is less effective in causing thyroid disease than are x-rays, and this might be especially true when the ^{131}I doses are spread out over several years (dose protraction tends to reduce the amount of cell damage that cells cannot repair).

Comment: Effects of ^{131}I and x-rays should be considered equivalent.

Comment: The ^{131}I dose-response relationship for thyroid disease is not linear. There is a threshold for radiation effects on the thyroid.

Response: Human data on thyroid cancer after gamma-ray exposures (in the Japanese atomic-bomb study) or medical studies of x-ray exposure are best fitted as a linear dose-response association (Ron and others, 1989), although a threshold at some low dose under 0.1 Gy (10 rad) cannot be conclusively ruled out. There are no compelling biologic reasons for the shape of the dose-response curve to differ greatly for ^{131}I exposure. In fact, the best study comparing the effect of x-ray and ^{131}I exposure in rats found essentially the

same dose-response relationships for both (Lee and others, 1982). Furthermore, the largest body of data on thyroid cancer after childhood exposure to ^{131}I—the Chernobyl data—also show a substantial excess of thyroid cancer in the estimated dose ranges of 0.1–0.5 Gy (10–50 rad) and 0.5–1.0 Gy (50–100 rad), although there is a suggestion that ^{131}I is only about 50% as effective as gamma rays in inducing thyroid cancer (Jacob and others, 1998).

APPENDIX C

ESTIMATED THYROID DOSES FROM "GLOBAL" WEAPONS TEST FALLOUT IN AREAS DOWNWIND FROM HANFORD

Report prepared for the NAS/NRC Board on Radiation Effects Research
Harold L. Beck
Formerly, Director, Environmental Science Division
USDOE Environmental Measurements Laboratory
April 1, 1999

I. INTRODUCTION

The CDC-FHCRC study considered only fallout from weapons tests at the Nevada test site as a potential confounder using doses estimated from the NCI study[1] on doses from I-131 from Nevada weapons tests. However, a concern was expressed by an NAS/NRC Committee that "global" fallout from weapons tests conducted outside the U.S. should also have been considered. The author of this report was asked to examine the literature and any available data and estimate the doses that may have been received by the population of counties downwind from Hanford from this "global" fallout. This report estimates the thyroid doses received by infants, children, teens, and adult males for each year of significant testing for three areas (Benton, Franklin and Adams Counties, Walla Walla County and Stevens County, WA) and compares the results with the NCI results for these same counties for NTS fallout.

Table 1 lists estimated fission yields for each month during the period 1952–63 when tests in the atmosphere were conducted at sites in the Northern Hemisphere other than the Nevada Test Site. The total number of tests was over 500. Additional tests were conducted by China in the 1970's, however, the additional yield, and thus additional fallout, was small compared to the fallout in the years shown and thus was not considered in this report. The fission yields shown represent the sum of the estimated yields from all tests conducted during the indicated month and are presented here only to indicate the months when significant fallout might have occurred over large areas of the world. The exact yields of many tests were not announced and the ratios of fission yield to total yield are classified but are estimated to be on average 50%. Thus the fission yields given in Table 1 should not be taken as highly accurate but rather as a good indication of the relative yields as a function of time. It can be noted, however, that the total of the estimated fission yields from all atmospheric tests is consistent with measurements of Sr-90 deposition.[2]

As can be seen from the Table, the total fission yield from these tests is over 150 MT compared to a total yield of about 1 MT from the approximately 80 tests carried out at the NTS. However, most of these tests were carried out at sites far from the U.S., primarily in the south Pacific at Bikini and Enewetok or in the Soviet Union. Furthermore, because most of the yield was from multi-megaton thermonuclear tests, it is estimated that about 80% of the debris were injected into the stratosphere. For the approximately 20% of the debris injected into the troposphere, a considerable fraction was probably deposited locally or regionally, particularly debris from surface shots as opposed to air bursts. The amount of debris reaching various areas of the U.S. also depends on the location of the test site. The tests in the Pacific were conducted at latitudes fairly close to the equator while the tests in the Soviet Union were at fairly high latitudes. Troposphere fallout clouds tend to travel around the globe remaining primarily in the same latitudinal band. Finally, the deposition at any particular site depends primarily on whether or not rain occurred at the time the debris was overhead. Thus it is not surprising that, as will be shown, fission yield is only a very crude indicator of the fallout deposition in the U.S.

Unfortunately, there is only a limited amount of actual data on fallout deposition at particular sites in the U.S., particularly for short-lived nuclides, and virtually none for I-131 deposition. Fortunately, however, data on actual I-131 concentrations is milk are available for some of the testing period.

II. AVAILABLE DATA

A number of potential data sources were reviewed including the Quarterly Fallout Reports of the USAEC Health and Safety Laboratory, Reports of Hearings conducted by the US Congress Joint Committee on Atomic Energy. Monthly reports issued by the Public Health Service, Environmental Monitoring Reports from the Hanford Site and selected literature sources. Data that was found that are relevant to this study are discussed below.

Milk:

I-131 was measured in milk from a number of dairy farms in the counties adjoining the Hanford site for 1961 and 1962. These data, taken from the 1961 and 1962 Site Environmental Monitoring Reports[3] and averaged by month, are given in Table 2. Note that the data indicate significant amounts of I-131 in milk in the summer of 1961, even though there were no weapons tests before September 1961! Note also that the levels in milk varied significantly from dairy to dairy, particularly in late 1962, probably reflecting differences in the amount of feed received from fresh pasture. Soldat[1] indicated that on average only about 50% of the cows were probably on pasture during October and November. The Hanford Environmental Monitoring Report for 1958[4] also contains some information on I-131 in milk. However, it appears only 4 measurements were made at a single dairy farm, Riverview. The average of these four measurements was about 150 pCi/L. No information was given on exactly when the measurements were made. The Public Health Service (PHS) starting in 1958 also measured I-131 in milk.[5] Only a limited number of sites (about 12) were sampled in 1958, and for many the sampling did not begin until July or August. The sampling network was expanded to about 60 sites in 1961 and 1962. Selected PHS data are presented in Table 3.

Deposition:

Unfortunately, no data on the deposition of I-131 is available, either for the Hanford area, or for other areas of the U.S. Soldat[1] presents a few values for the I-131 concentration in forage during 1961 and 1962 from which one can infer the approximate deposition. However, beginning in 1957, the USAEC's Health and Safety Laboratory (HASL) began measuring Sr-89 deposition along with Sr-90 using pots and ion exchange columns.[6] Although Sr-89 has a half-life of about 50 d as opposed to the 8 days for I-131, it should still be a useful surrogate for estimating the deposition of other short-lived radionuclides injected into the

troposphere. Unfortunately, Sr-89 was not measured near Hanford. However, as shown in Table 4, the total annual Sr-89 deposition rate per cm of precipitation appears to be fairly constant over large regions of the country in any given year. Although, the data in the table are for the entire year for months when fallout occurred, data for months of heavy fallout exhibit the same general pattern. (Note that one cannot compare differences from year to year in Table 4 since the fallout occurred over different intervals). This is not unexpected since the debris clouds would be expected to have dispersed considerably by the time they reached the U.S. and it is well known that the primary mechanism for fallout far from the immediate test sites is from precipitation scavenging. Thus it should be possible to infer the deposition of Sr-89 near Hanford from these data and the monthly precipitation values from the counties near Hanford.

A limited amount of data on Ba-140 (Half-life = 12 d) was available for three sites; Pittsburgh, Westwood, NJ, and Richfield, CA.[7] These data were used to corroborate the estimated relationship between Sr-89 and I-131 deposition.

Finally, for years prior to 1957, the only deposition data available was from measurements from the HASL Gummed-Film Network. Unfortunately, the gummed-film data for 1954, 1956 and 1958 have not been re-analyzed as were the data for the years of NTS testing.[8] However, Harley[9] presented estimates of gamma dose made using the raw gummed-film measurements. While the absolute values are probably suspect[7], the relative annual estimates should still provide a reasonable estimate of the relative short-lived radionuclide deposition that can be used to estimate the fallout deposition in 1956 and 1954.

Precipitation:

The monthly precipitation for Benton, Franklin Adams, Walla Walla, and Stevens Counties taken from historical US Weather Service Records are listed in Table 5. The data for Benton, Franklin and Adams were averaged since the variations in monthly precipitation in these three counties were small. The

monthly precipitation for Yakima and Kittitas Counties, which supply milk to the Hanford area, were similar to that for Adams, Franklin and Benton Counties.

I-131 Releases from Hanford:

References 1 and 2 also provide data on the I-131 released from the Hanford stacks in 1958, 1961 and 1962. The average daily releases were 1.2 Ci/d in 1958, 0.54 Ci/d in 1961 and 0.33 Ci/d in 1962. This information was used to estimate the Hanford contributions to the activity in milk values presented in Table 2.

III. METHODOLOGY

The basic methodology used to estimate the doses from I-131 near Hanford was as follows. For 1961 and 1962, the doses for Franklin, Adams and Benton counties were estimated directly from the measured I-131 in milk at farms near Hanford after first correcting the milk data for Hanford plant contributions. Since all three counties are part of the same milk shed and receive similar amounts of rain, the estimated milk concentration data was assumed to apply to all three counties and a single set of dose estimates was made for these three counties. (The doses for Yakima and Kittitas Counties, which are also part of the same milk producing area, would also be similar to those for Franklin, Adams and Benton due to the similar rainfall pattern). The Sr-89 depositions for 1957, 58, 61 and 62 were then estimated from the Sr-89 depositions at sites in the western U.S. and the measured monthly rainfall for Hanford area counties. The ratios of the deposition for Walla Walla county and Stevens County to Benton-Franklin-Adams were used to estimate the milk concentrations for those counties for 1961 and 1962 from the measured milk near Hanford. The calculated deposition in 1958 relative to 1961 and 1962 was then used to estimate the relative concentrations in milk for 1957 and 1958 for all three county areas. The Sr-89 deposition for 1956 and 1954 relative to 1958 was estimated from the

gummed-film data and the concentrations in milk were assumed to vary the same as the estimated deposition. Finally, the thyroid doses for each county were estimated using the same conversions factors from milk concentration to dose for that county used in the NCI study. Further details are discussed below.

Estimated Fallout I-131 in Milk for Benton, Adams, Franklin Counties:

The average annual I-131 concentration in milk for the three dairies sampled was 48 pCi/L in 1961 (see Table 2). However, the average annual concentration based only on data during months of global fallout is 33 pCi/L. Based on the activities measured during the summer months prior to weapons testing, it is estimated that about 40% or 19 pCi/L of the average annual activity in milk during months with fallout was from I-131 released from Hanford. The Hanford Plant contribution probably varied from month to month depending on local meteorological conditions so that the estimated plant contribution is somewhat uncertain. The measured concentrations were highest for the Ringold dairy farm, reflecting a probably greater fraction of feed from fresh pasture, particularly during the fall months when fallout occurred. Therefore, the average annual concentration for milk from Ringold, reduced by 40% ($57 \times 0.365 = 21 \times 0.6 =$ 12 nCi-d/L), was adopted as the best estimate of the concentration in milk for 1961 for cows on pasture.

For 1962, again, adopting the data from the Ringold farm as most representative of cows on maximum pasture, the total activity measured was 75 pCi/L or about 15% higher than in 1961. As discussed earlier, the I-131 releases from Hanford in 1962 were 0.33 Ci/d versus 0.54 Ci/d in 1961. Thus, on average one would expect a Hanford contribution of about 60% that for 1961 or about 20% of the activity measured in Ringold milk in 1962. The concentration of I-131 in milk for 1962 for cows on pasture was thus estimated to be $0.8 \times 75 \times 0.365 = 22$ nCi-d/L.

For 1958, only 4 measurements were reported, all for the Riverview farm. The average of 150 pCi/L would correspond

to a milk concentration of about 55 nCi-d/L. However, the releases from Hanford during 1958 averaged 1.2 Ci/d versus 0.54 Ci/d for 1961 implying a contribution from Hanford of about 15 nCi-d/L based on the 1961 estimate. The net contribution of 40 nCi-d/L, although highly uncertain, is in reasonable agreement with the estimate of 30 nCi-d/L adopted for this study based on relative Sr-89 deposition. (See next paragraph.)

Sr-89 Deposition for Benton-Adams-Franklin, Walla Walla and Stevens Counties:

The Sr-90 deposition in each of the three sets of counties was estimated on a monthly basis by multiplying the monthly precipitation listed in Table 5 by a weighted average of the Sr-89 deposition per cm of rain???[6] at the following western U.S. sites where Sr-89 was measured: Seattle, Medford, Salt Lake City, Vermillion, Richfield, weighting by the inverse of the distance from each site. Only data for days with rain were used. The monthly estimates were then summed to provide an estimate of total annual Sr-89 deposition. For 1957, it was estimated that about 1/3 of the estimated Sr-90 deposition at the above western U.S. sites where Sr-89 was measured resulted from tests at the NTS during August and September and this contribution was not included. Also, the total deposition in 1962 through January 1963 was calculated but it was decided to use only the deposition through November 1962 for calculating milk concentrations. Most of the additional fallout occurred in late December and January when cows were not on pasture from high yield tests conducted in late December.

I-131/Sr-89 Deposition:

It is generally accepted that the average residence time for fallout released into the troposphere is about 30d. It can then be shown that on average about 21% of the I-131 and 63% of the Sr-89 released into the troposphere will deposit before decay. Since the fission yield of I-131 is about 7 × that of Sr-89, one would thus expect on average an I-131/Sr-89 ratio of about 2.3. However, as mentioned previously, about 80% of the fission products were

probably injected into the stratosphere where the average removal half time is about 1 to 1.5 years. One would then expect only about 10% of the Sr-89 injected into the stratosphere (i.e. about 8% of the total produced) to be deposited before decay. Almost all of the I-131 injected into the stratosphere would decay before being deposited. Thus the actual ratio of Sr-89 to I-131 expected on average to be about $(0.21 \times 0.2) \times 7 / (0.63 \times 0.2 + 0.1 \times 0.8) = 1.5$. This is of course a very rough approximation. The actual ratio will vary from test to test and site to site. The ratio at any given time will depend on the amount of debris injected into the stratosphere and variations in stratospheric residence time from season to season. It will also depend on variations in Sr-89 to I-131 deposition from site to site due to the fact that the I-131 is deposited over a period of less than a month while the Sr-89 is deposited over several months. However, since the annual Sr-89 estimates generally reflect the sum of fallout from a large number of tests and seasons, a ratio of about 1.5 would be expected to reasonably reflect the annual average I-131/Sr-89. Furthermore, the limited data on Ba-140 deposition tends to confirm the estimate of 1.5 as being a reasonable average. Ba-140 was measured at Westwood, NJ and Richmond, CA during 1961. The average ratio over several months of data was 2.5. Ba-140/Sr-89 measurements at Pittsburgh and Richmond during 1958, again for several months of data, averaged 2.3. Since Ba-140 has a half life of 12 d versus 8 d for I-131, the expected Ba-140/Sr-89 ratio would be about 1.5 times that of I-131, implying an I-131/Sr-89 ratio of 2.4/1.5=1.6. Thus the Sr-89 deposition estimates were multiplied by 1.5 to provide a rough estimate of the annual I-131 deposition. As will be shown later, these I-131 deposition estimates, are in reasonable agreement with the milk estimates based on comparable data from NTS. However, it is important to note that the absolute I-131 deposition estimates were not used to estimate doses. Only the relative Sr-89 depositions from year to year were used to estimate doses.

Relative Deposition in 1954 and 1956 Versus 1958 from Gummed-Film:

The data in Table 6 indicate that the relative deposition in 1956 relative to 1958 was about 0.4-0.6, about in the same ratio as the fission yields. However, the gummed-film data for 1954 indicate a ratio of only about 0.2-0.4 that of 1958, much less than the relative fission yield. This is not exactly unexpected, however, since all of the tests conducted in 1954 were surface shots compared to only about 2/3 of the yield in 1958 being from surface shots in 1958 and 3/4 in 1956. Surface shots would result in a much larger proportion of the debris being deposited locally and regionally as opposed to globally. Since these particular gummed-film data represent very crude estimates of short-lived fallout, and the data in Table 6 does not reflect variations from site to site due to variations in precipitation from year to year nor corrections for differences in stratospheric deposition from year to year that are known to be included in the annual estimates, it was decided to adopt a ratio of deposition for each year of 0.4 of the 1958 deposition, even though the value may be somewhat conservative for 1954. An improved estimate of the deposition for these years might be possible with a re-evaluation of the gummed-film data as was done for the years of NTS testing, however, that was beyond the scope of the present assessment. At any rate, the uncertainty in overall deposition estimates for 1954 and 1956 is probably still no worse than a factor of 2-3, comparable to that for NTS deposition in these counties.

I-131 Milk Concentrations for 1954, 1956, 1957 and 1958:

The milk concentrations for each county(s) for 1958 were estimated from the 1961 milk concentrations based on the relative deposition in 1958 versus 1962. However, since about 10% of the deposition in 1958 occurred in February and March, before the pasture season, the 1958 milk concentrations were reduced by 10% below this ratio. For other years, almost all the fallout occurred during the pasture season and thus the milk concentrations were assumed to vary directly as the estimated

depositions. The assumption that the concentration in milk will vary in direct proportion to the deposition is not strictly valid since the interception of fallout by vegetation depends on the rainfall rate and fraction of feed from fresh pasture. However, over an entire pasture season encompassing many fallout events, the variations should average out and the approximation should be fairly reasonable. The relative milk concentrations inferred for 1958 versus 1961 and 1962 are consistent with the ratios of the concentration in milk at sites in the western U.S. measured by the PHS (see Table 3). The relatively small variation in the PHS milk values over large regions is also consistent with the deposition estimates based on the Sr-89 data. The absolute concentrations from the PHS network sites are generally lower than those estimated for the Hanford area. This probably reflects the fact that each PHS measurement is an average over a large milkshed that incorporated cows that were not always on fresh pasture while the Hanford area estimates are based on a maximum fresh pasture scenario.

Dose Calculation:

The doses were calculated from the milk concentrations using average milk to dose conversions for each age group for each county inferred from the NTS doses (see Table 7). The NTS dose per unit annual milk concentration for infants and children vary from year to year (Table 7) reflecting the fact that the age grouping will change depending on the exact dates of the fallout. However, the present estimates use an average value. This may have introduced a small bias into the estimated doses for infants and children from "global" fallout.

IV: RESULTS

Table 7 provides the estimated "global" fallout doses calculated as described above for the three regions downwind from Hanford. Benton, Adams and Franklin counties were combined since the deposition was essentially the same in these three

counties. The NTS dose estimates are actually the estimates for Franklin County. However, the NTS dose estimates for all three counties are very similar. Walla Walla was considered separately because of its significantly higher rainfall rate and much higher NTS fallout while Stevens County is included to indicate the doses further away from the Hanford site. Stevens County probably also can be considered to have similar deposition as Spokane County and probably supplied much of the milk for Spokane residents???[1]. The results shown in Table 7 indicate that the "global" fallout doses near Hanford were lower but of the same order as the doses from NTS fallout.

Since the dose estimates given here are essentially ratios of the measured milk concentrations of fresh farm milk when cows are on pasture, they were compared with NTS estimated doses for milk from a backyard cow. These doses are similar to those for fresh milk consumed on the farm. For other classes of milk drinker, the relative global to NTS doses would be similar although the absolute doses would be lower. Note that the estimated doses in Table 7 are doses from ingestion of milk and do not include the small additional doses from other foods.

The absolute I-131 deposition estimates given in Table 7 were not used to calculate doses. Only the relative depositions from year to year were used. It is, however, encouraging to note that the estimated I-131 depositions from global fallout relative to the estimated milk concentrations are in concordance with the same ratios for NTS fallout where the milk concentrations were calculated directly from the estimated deposition. This indicates a measure of self-consistency that supports the validity of the methodology used to estimate the milk concentrations and resulting doses. As discussed previously, the exact ratio of milk concentration to deposition will vary somewhat due to variations in the amount of feed from fresh pasture and variations in the interception of fallout by vegetation. The estimated I-131 depositions for 1961 and 1962 are also in reasonable agreement with the limited data on I-131 concentration in fresh forage

reported by Soldat in reference 3, using mass interception factors reported in reference 1.

The I-131 thyroid doses from NTS fallout were estimated to have uncertainties of about a factor of 3–4. Most of this uncertainty was from the estimate of fallout deposition, which was based on interpolation of the sparse data from the gummed-film network. The "global" fallout estimates for 1961 and 1962, being based on actual data are probably less uncertain, probably no worse than a factor of 2. The doses for 1958, based on the relative interpolated Sr-89 deposition are also probably less uncertain than the NTS with the 1958 to 1961/1962 ratios being accurate to about +/- 50%. However the uncertainty in dose estimates for 1958 is of course correlated to the uncertainty in the 1961, 1962 milk data and assumes the same ratio for deposition to milk concentration. Finally, the 1956 and 1954 estimates are more uncertain, perhaps as much as an additional factor of 2. Since the same factors were used to convert from milk concentration to dose for both NTS and global fallout, any error in this conversion (other than as discussed for infants and children) would be about the same for both fallout sources. Thus, one concludes that considering the uncertainty in both the NTS and "global" fallout dose estimates, the differences between the two sets of dose estimates are probably not statistically significant. It is interesting to note also that on the basis of the observed milk data from Hanford, the doses to the population around Hanford from site releases were a significant fraction of those from fallout in these same years.

The dose estimates in Table 7 may be compared to the population-weighted estimate for the 50–60 degree latitude band of the Northern Hemisphere estimated by UNSCEAR[2] of 1.6 mSv from all weapons fallout. This estimate is of course a very rough average, which assumes uniform deposition of I-131 over the entire latitude band. The Hanford area would be expected to have lower doses than the average due to the significantly lower average precipitation, but higher due to being relatively closer to the test sites in the Pacific.

Finally, it should be kept in mind when comparing these to doses from the Hanford HEDR Study, that persons exposed as infants in 1946, i.e. those with the highest doses, would have been exposed to NTS and global fallout as teenagers. Infants exposed to global fallout would of course not have been exposed during the major releases from Hanford.

Table 1. ESTIMATED FISSION YIELD*

Month	MT	Month	MT
Nov-52	**0.9**	Feb-58	1.4
		Mar-58	0.4
Mar-54	8.7	Apr-58	0.0
Apr-54	3.6	May-58	1.2
May-54	5.3	Jun-58	5.7
Jun-54	0.0	Jul-58	5.8
Jul-54	0.1	Aug-58	4.3
1954 Total	**18**	Sep-58	1.1
		Oct-58	5.8
May-56	3.7	Nov-58	0.0
Jun-56	0.9	**1958**	**26**
Jul-56	6.5		
Aug-56	0.5	Sep-61	3.0
Sep-56	0.6	Oct-61	6.0
Oct-56	0.0	Nov-61	11.0
Nov-56	0.0	Dec-61	0.0
Dec-56	0.4	**1961 Total**	**20**
1956 Total	**13**		
		May-62	2.0
Sep-57	0.4	Jun-62	3.0
Oct-57	2.3	Jul-62	8.0
Nov-57	0.6	Aug-62	8.0
1957 Total	3	Sep-62	9.0
		Oct-62	20.0
		Nov-62	13.0
		Dec-62	11.0
		1962 Total	**74**

* NTS tests not included
Source: Reference 2 and unpublished data.

Table 2. Hanford Milk Measurements (pCi/L, average for month)

		Ringold	Riverview	Benton	Eltopia	Mesa	mean
J	1961	28	<50	<50	ND	ND	<50
F		17	<50	<50	ND	ND	0
M		<50	<50	<50	ND	ND	17
A		33	43	<50	ND	ND	27
M		42	37	25	ND	ND	33
J		52	30	21	ND	ND	36
J		<50	ND	<50	ND	ND	<50
A		<50	<50	<50	ND	ND	<50
S		47	58	<50	ND	ND	40
O		357	67	120	ND	ND	180
N		34	98	312	ND	ND	150
D		16	9	13	ND	ND	17
Annual Average		**57**	**37**	**48**	ND	ND	**48**
Estimated Hanford Contribution		**23**	**15**	**19**			**19**
J	1962	5	3	2	ND	ND	3
F		4	3	1	1	1	3
M		6	2	2	1	2	3
A		30	3	1	12	8	10
M		4	3	1	2	2	3
J		29	23	23	4	3	16
J		4	7	3	5	6	5
A		12	14	7	17	12	12
S		109	58	52	72	17	62
O		76	38	43	113	177	90
N		344	32	31	261	283	190
D		277	11	23	10	31	71
Annual Average		**75**	**16**	**16**	**41**	**45**	**39**
Estimated Hanford Contribution		**15**	**3**	**3**	**8**	**9**	**7**

Table 3. PHS Milk Data, Average Annual Concentration (pCi/L)

	1958	1961	1962	58/61	58/62	62/61
Fargo	>25(45E)	19	23	>1.3(2.3 E)	>1.1	1.2
Sacramento	31	6	13	5.2	2.4	2.2
Spokane	>16(40E)	11	53	>1.5(3.6E)	>0.3	4.9
SLC	31					
Atlanta	>14(20E)	11	22	>1.3		
Chicago	>22(30E)	28	36	>0.8		1.3
NYC	28	24	30	1.2	0.9	1.3
Seattle		>20	26			<1.3
Portland		22	25			1.2
Helena		>24	36			<1.5
Cincinnati	33					
Network Avg.	**35E**	**21**	**31**	**1.7E**	**1.2E**	**1.5**
Hanford	80E	34	60	2.5E	1.3E	1.8

E = Estimated (data available for only part of year).

Table 4. Sr-89 Deposition per cm rain (nCi/m2 per cm*)

	1958		1961		1962	
Site	Precip	Sr-89	Precip	Sr-89	Precip	Sr-89
New York	85	1.3	28	2.4	78	2.0
Pittsburgh	77	1.3	25	1.6	63	2.1
Chicago	60	1.3	54	>1	42	2.0
Vermilion, SD	33	2.5	14	5.7	57	2.2
Salt Lake City	18	2.9	12	5.4	20	3.7
Medford, OR		ND	24	0.9	50	1.9
Richmond, CA	18	2.1	17	2.0	51	2.0
Seattle	43	2.5	20	2.5	59	2.5

* Totals for months with fallout. Precipitation in cm.

Table 5. Monthly Precipitation (cm)

Year	Month	Adams, Benton, Franklin Counties	Walla Walla County	Stevens County	Seattle
1952	Nov-52	0.8	1.1	1.8	4.2
1954	Mar-54	2.0	2.8	2.6	5.4
	Apr-54	0.8	3.4	2.3	6.9
	May-54	1.0	1.7	3.5	4.5
	Jun-54	1.1	3.0	3.3	4.6
	Jul-54	3.0	3.3	5.0	10
	Total	**8**	**14**	**17**	**31**
1956	May-56	2.1	6.3	2.7	1.7
	Jun-56	2.3	2.4	3.3	7.1
	Jul-56	0.7	0.3	2.0	0.2
	Aug-56	0.9	3.8	3.3	2.4
	Sep-56	0.4	0.3	0.4	5.6
	Oct-56	2.6	5.4	4.9	10.4
	Nov-56	0.6	1.7	0.8	4.1
	Dec-56	1.5	4.9	2.5	6.8
	Total	**11**	**25**	**20**	**38**
1957	Sep-57	0.4	4.0	2.0	4.7
	Oct-57	1.3	4.0	4.0	9.2
	Nov-57	2.0	5.3	7.0	12
	Total	**4**	**13**	**13**	**26**
1958	Feb-58	4.8	5.0	9.0	14.3
	Mar-58	2.0	4.3	3.2	6.6
	Apr-58	2.8	9.1	7.1	3.7
	May-58	1.7	5.5	1.6	2.3
	Jun-58	1.2	2.5	3.5	2.1
	Jul-58	0.3	0.0	3.7	0
	Aug-58	0.2	0.0	0.6	1.1
	Sep-58	0.3	1.6	2.0	3.7
	Oct-58	0.6	1.3	2.5	7.9

Year	Month	Adams, Benton, Franklin Counties	Walla Walla County	Stevens County	Seattle
	Nov-58	2.8	5.5	9.6	15.9
	Total	**17**	**35**	**43**	**58**
1961	Sep-61	0.5	0.4	0.8	1.6
	Oct-61	0.7	3.4	3.8	7.2
	Nov-61	2.0	4.6	4.7	11.2
	Dec-61	2.5	6.2	10.0	14.2
	Total	**6**	**15**	**19**	**34**
1962	May-62	4.5	10.5	5.7	2.7
	Jun-62	0.4	0.7	2.7	1.5
	Jul-62	0.0	0.0	0.1	4.0
	Aug-62	1.1	1.1	2.0	5.1
	Sep-62	1.1	5.2	3.2	8.7
	Oct-62	3.2	8.5	5.9	18.1
	Nov-62	2.2	5.9	7.6	9.7
	Dec-62	2.2	6.9	5.2	4.9
1963	Jan-63	1.3	2.3	0.6	4.5
	Total	**16**	**41**	**33**	**59**

Table 6. Ratio of Gummed-Film Gamma Dose Estimates

Site	1954/58	1956/58
Boise, ID	0.3	0.5
Billings, MT	0.3	0.7
Salt Lake City	0.4	0.5
Grand Junction, CO	0.5	0.6
Seattle, WA	0.3	0.4
Medford, OR	0.1	0.4
San Francisco, CA	0.2	0.4
Yield ratio	0.7	0.4

Yield from surface shots: 100% in 1954, 75% in 1956, 67% in 1958.

Table 7. Estimated Doses by Age-Group and County

Benton, Adams, Franklin							
		I-131 Deposition nCi/m^2	Milk nCi-d/L	Infant† mSv	Child† mSv	Teen† mSv	Adult mSv
G	1954	20	13	3	1.5	0.52	0.20
L	1956	20	13	3	1.5	0.52	0.20
O	1957	11	6	1	0.7	0.24	0.09
B	1958	50	30	6	3.3	1.20	0.45
A	1961	18	12	2	1.3	0.48	0.18
L	*1962	36 (63)	22	4	2.4	0.88	0.33
	TOTAL	**155**	**96**				**1.5**
N	1952	115	81	15	8.3	3.1	1.2
T	1953	53	29	6	3.4	1.2	0.5
S	1955	40	29	6	3.6	1.3	0.5
	1957	76	83	16	8.7	3.2	1.2
	TOTAL (Franklin)	284	222				**3.3**

Walla Walla							
		I-131 Deposition nCi/m^2	Milk nCi-d/L	Infant† mSv	Child† mSv	Teen† mSv	Adult mSv
G	1954	36	24	5	2.9	1.0	0.38
L	1956	36	24	5	2.9	1.0	0.38
O	1957	30	18	4	2.2	0.72	0.29
B	1958	90	53	12	6.4	2.1	0.84
A	1961	33	22	5	2.6	0.88	0.35
L	*1962	124 (190)	75	17	9.0	3.0	1.2
	TOTAL	**350**	**216**				**3.4**
N	1952	175	95	18	11	3.8	1.4
T	1953	80	46	10	5.5	2.0	0.8
S	1955	60	36	9	4.9	1.8	0.7
	1957	800	540	94	51	19	7.1
	TOTAL	**1120**	**717**				**10**

Stevens		I-131 Deposition nCi/m^2	Milk nCi-d/L	Infant[†]mSv	Child[†] mSv	Teen[†] mSv	Adult msv
G	1954	60	45	7	4.9	1.8	0.58
L	1956	60	45	7	4.9	1.8	0.58
O	1957	30	20	3	2.2	0.8	0.26
B	1958	150	90	14	9.9	3.6	1.2
A	1961	45	30	5	3.3	1.2	0.39
L	[*]1962	106(145)	66	11	7.2	2.6	0.86
	TOTAL	**450**	**296**				**3.8**
N	1952	150	110	18	10	4.1	1.3
T	1953	250	150	29	16	6.4	2.1
S	1955	50	47	6	5	2.1	0.5
	1957	90	100	15	9	4.0	1.3
	TOTAL	**550**	**460**				**5.2**

1951, 1958 NTS, 1952 global fallout doses negligible.

[*] deposition through November (deposition through Jan 1963 in parenthesis).

[†] Totals not applicable because individuals will change age-category.

Infant: 1–5 months

Child: 1–4 yr.

Teen: 10–14 yr.

REFERENCES

[1] Estimated Exposures and Thyroid Doses Received by the American People from Iodine-131 in Fallout Following Nevada Atmospheric Nuclear Bomb Tests. National Cancer Institute, 1997.

[2] Reports of the United Nations Scientific Committee on the Effects of Atomic Radiation. UNSCEAR, 1982, 1993.

[3] Hanford Environmental Monitoring Program Annual Report for 1961, HW-71999; Hanford Environmental Monitoring Program Annual Report for 1962, HW-76526. See also J.K. Soldat, The Relationship between I-131 Concentrations in Various Environmental Samples, Health Physics 9, 1167, 1963.

[4] Anderson, B.V., Hanford Environmental Monitoring Annual Report-1958, HW-61676.

[5] Radiological Health Data. U.S. Public Health Service. Monthly Reports for 1958, 1961, 1962.

[6] Final Tabulation of Monthly Sr-90 Fallout Data: 1954–1976. USERDA Report HASL-329, 1977.

[7] USAEC Health and Safety Laboratory Quarterly Fallout Reports for 1958, 1959, 1961, 1962.

[8] Beck, H.L., Helfer, I.K., Bouville, A, Dreicer, M. Estimates of Fallout in the Continental U.S. from Nevada Weapons Testing Based on Gummed-film Monitoring Data, Health Physics 59(5), 565–576, 1990.

[9] Harley, J.H., N.A. Hallden and S.Y. Ong. Summary of Gummed-film Results Through December 1959. U.S.A.E.C Report HASL-93, Sept. 1960.

Appendix D

Equation Relating to the Effect of Error in Assessing Childhood Milk-consumption

The approximation

$$\frac{N}{1-M+MR^2}$$

for the inflation in sample sizes due to errors in milk consumption is derived by noting that if the dose estimate is linear both in milk intake, I_m, and also in other variables, *W,* which are independent of milk consumption, then the variance of the dose distribution can be divided into two portions. For example if dose, *D,* is approximately equal to

$$D = a + bW + cI_m$$

Then the variance of the dose distribution is approximately equal to

$$\text{var}(D) = b^2 \text{ var}(W) = c^2 \text{ Var}(I_m).$$

Now if milk consumption is estimated using a questionnaire to give value Q_m then which has correlation R with true milk consumption, then in order to estimate a linear dose response function in dose with no dose error attenuation bias, we replace D in the response model with

$$E(D \mid Q_m) = a + bE(W \mid Q_m) + cE(I_m \mid Q_m)$$

(Carroll and others, 1995). The variance of this quantity determines the sample size required to detect a nonzero dose response, as described in chapter X on statistical power. Assuming

also that the non-milk variables are independent of the questionnaire estimate of milk consumption, we have

$$\text{var}(E(D \mid Q_m)) = b^2\ \text{var}(W \mid Q_m) + c^2\ \text{var}(_m \mid Q_m) = b^2\ \text{var}(W) + c^2 R^2\ \text{var}(I_m).$$

Letting $M = c^2 \text{Var}(I_m / \text{Var}(D)$ be the fraction of the variance of dose that depends upon milk and *1-M* the remainder, the we have

$$\text{var}(E(D \mid Q_m) = \text{var}(D)(1\text{-}M + MR^2).$$

Thus if we do not know true milk consumption, but only Q^m, in order to have the same power to detect a linear dose response function we must (from the equation for the noncentrality parameter) have

$$N^2\ \text{var}(D \mid Q^m) = N\ \text{var}(D)$$

where N is the sample size needed with no errors in estimating milk consumption, and N^2 is the sample size required when Q^m is used to estimate milk. Therefore the inflation in sample size needed is

$$N_2 = \frac{N}{1 - M + MR^2}.$$

Glossary

Acellular or hypocellular specimen: A specimen that contains no cells or has fewer cells than normal, such as when using a fine-needle aspiration.

Activity: The amount of a radioactive nuclide in a particular energy state at a given time. Units of activity are becquerel (Bq) and curie (Ci); 1 Ci = 3.7×10^{10} disintegrations per second, and 1 Bq = 1 disintegration per second.

Acute release: An instantaneous release into the environment, similar to a puff of radioactivity.

Ascertainment bias: Systematic error due to a difference in characteristics between those who choose to participate in a study and those who do not.

Attenuation: In statistics, a bias in a study in which the estimated strength of an observed dose response tends to be lower than the true dose response. Attenuation is expected in a dose-response study in which exposure is estimated with error under the classical measurement-error model.

Berkson measurement error: A statistical model of errors in exposure estimates in which the true value for exposure of a study participant is assumed to be randomly distributed symmetrically around the estimate of exposure of that individual. Unlike classical measurement error, Berkson error does not reduce the apparent slope of the dose association between an exposure and an outcome variable.

Blinded interviews or examinations: A study design in which the interviewer or examiner does not know which group the subject belongs to. The intent of blinding procedures is to eliminate the potential for even unconscious biases and prejudices of the interviewer or the examiner to affect the results.

CIDER model: A computer program used for calculation of individual doses from environmental radionuclides. The model estimates doses from four pathways: submersion in contaminated air, inhalation of contaminated air, irradiation from contaminated surfaces, and ingestion of contaminated farm products and vegetation.

Classical measurement error: A statistical model of errors in exposure estimates in which the exposure estimate for a study participant is assumed to be randomly distributed symmetrically around the true value of exposure of that individual. This type of measurement error reduces the apparent strength of the association between an exposure and an outcome variable.

Cohort: The identified subset of a defined population with various degrees of exposure to factors that the investigator wants to study.

Computer realizations:	Outcomes of a model in which different parameter values are chosen randomly from specified distributions of values.
Confidence interval:	A range of values of a variable of interest, such as an effect of dose on disease risk, constructed to have a specified probability (typically 95%) of including the true value of the variable.
Confounder (or confounding variable):	A factor that distorts the magnitude of the effect of a study factor on disease risk. A confounding factor is a determinant of the disease risk and is unequally distributed among those with low and high exposures.
Congenital anomalies:	Birth defects that occur after fertilization of the embryo and result from developmental errors as the embryo grows.
Correlation:	Most generally, the degree to which one phenomenon or random variable is associated with or can be predicted from another. In statistics, usually refers to the degree to which a predictive relationship between random variables exists. Correlation may be positive (both variables increase or decrease together) or negative or inverse (one variable increases when the other decreases).
Cytopathology:	The study of cellular disease, such as a microscopic examination that provides a characterization of stained cells or tissues from biopsied specimens.

Default value: If for some individuals no information is available on some factor (such as amount of milk drunk) used in a model to estimate exposure, some assumed average or typical value is assigned: the default value. In effect, this is the best guess of what the real value might be.

Dose: Quantity of radiation or energy absorbed per unit of mass. For special purposes, it must be appropriately qualified. If unqualified, it refers to absorbed dose. The unit of absorbed dose in this report is the gray (Gy); 1 Gy equals 1 joule of energy absorbed per kg of material, such as tissue. A subdivision of the unit is the milligray (mGy); 1 Mgy equals 0.001 Gy. One gray is equivalent to 100 rad.

Dose measurement error: Errors made in using estimates, such as those from dose-reconstruction methods, to predict true exposure of an individual; more-or-less random errors made because many things about that individual's exposure history are unknown, not mistakes in the calculation.

Dose misclassification: Errors in estimated dose or exposure.

Dose reconstruction: Process of estimating doses from past releases of radionuclides or chemicals to the environment.

Dose-response curve or dose-response relationship: A graph to show the relation between the dose of an exposure, such as radiation, and the degree of response or increase in a defined effect that it produces.

Dose-response regression coefficient: The regression coefficient in a model of dose response specifies the strength or shape of a dose-response curve.

Effect modifier: A factor that modifies the disease risk posed by the exposure of interest. For instance, radiation causes more thyroid cancer if the thyroid is exposed to 1 Gy in early childhood than if it is exposed to 1 Gy in adolescence or adulthood. Thus, age is an effect modifier.

Embargoing: A technique whereby an organization or scientific journal sends material to reporters in advance of an official release date. Reporters who receive the material agree not to publish an article about the material until it is officially released.

Epidemiologic studies: Studies designed to examine associations—commonly, hypothesized causal relations. They are usually concerned with identifying or measuring the effects of risk factors or exposures. The common types of epidemiologic studies are case-control studies, cohort studies, and cross-sectional studies.

Epidemiology: The study of the distribution and determinants of health-related states and events in populations.

Excess absolute risk: The increase in risk of disease that is related to exposure to a specified dose, or the arithmetic difference in risk of disease between exposed and unexposed subjects. Usually expressed as increase in risk per unit dose. See *Excess relative risk*.

Excess relative risk: The increase in relative risk of disease due to exposure to a specified dose. The mathematical distinction between excess absolute risk (calculated by simple subtraction) and excess relative risk is that the latter is calculated by dividing the risk of disease among exposed subjects by the risk among the unexposed and then subtracting 1.

Fallout: The radionuclides that become airborne after an environmental release and are then deposited on the ground.

Fine-needle aspiration: A procedure in which a fine, hollow needle is inserted into tissue to extract a small amount of tissue for microscopic evaluation.

Fractionation of exposure: One of the terms used to describe how an exposure was delivered over time. Exposures can be either single (brief), repeated (fractionated), or continuous (chronic).

Geostratum (plural, geostrata): A term used by the HTDS investigators to refer to the geographic areas (counties and cities) where the subjects in the study were born. The specific geostrata they used were Benton, Adams, Franklin, Walla, Okanogan, and Ferry-Stevens counties and the cities of Richland, Pasco of Kennewick, and Walla.

Heterogeneity: The presence in the population of subgroups that have different characteristics or different rates of disease, often for reasons that are not well understood.

Histogram:	A bar chart representing a frequency distribution; heights of the bars represent observed frequencies.
Hyperparathyroidism:	Disorder that is characterized by the excessive production of parathyroid hormones.
Hyperthyroidism:	Disorder that is characterized by the excessive production of thyroid hormones.
^{131}I dose coefficient:	The absorbed dose (in the thyroid) from unit intake of ^{131}I.
Incidence:	A measure of the rate at which cases of a disease occur in the population.
Isotope:	A variant of an element with the same number of protons in the nucleus but different numbers of neutrons. Some isotopes of an element are radioactive (radionuclides), and others are nonradioactive (stable nuclides).
^{85}Kr:	The isotope of krypton with a total number of protons and neutrons of 85.
Likelihood function:	Functions constructed from a statistical model and a set of observed data that give the probability of those data for various values of model parameters, such as regression coefficients in a dose-response curve. Parameter values that maximize the probability are the maximal-likelihood estimates of the parameters.

Likelihood-ratio method: A statistical method for assessing whether observed data from a study are consistent with a null hypothesis. The likelihood ratio is computed by dividing the value of a likelihood function that is calculated with maximal-likelihood estimates (see *Likelihood function*) by the value of the same likelihood function that is calculated with parameter values specified by the null hypothesis. The larger the likelihood ratio, the more strongly the study rejects the null hypothesis.

Linear regression: Statistical technique in which the value of a parameter (such as disease risk) for a given value of a factor x (such as dose) is assumed to be a + bx, where a and b are coefficients to be estimated from the data.

Linear-regression dose-response model: A dose-response curve in which the relationship between disease and exposure is specified by a linear regression.

Model coefficient: A parameter value used in a mathematical model.

Monte Carlo method: A technique for numerically approximating the solution of a mathematical or statistical problem by producing a distribution of some random variable, often generated by a computer. The name alludes to the randomness that is characteristic of the games of chance played in the gambling casinos in Monte Carlo.

Morbidity: A diseased condition or state; the incidence of a disease or of all diseases in a population.

Negative results: Results of a study that are consistent with a null hypothesis of no effect of exposure on disease risk.

Nonlocation, noncontact: Refer to the percentage of potential study subjects who could not be located or contacted, respectively.

Null hypothesis: A hypothesis that an exposure has no effect on disease risk. More generally, a null hypothesis typically specifies an assumption that any observed difference (in disease risk) between samples of a statistical population is accidental and not due to systematic causes (such as differences in exposure).

Null results: See *Negative results*.

Parity: The number of full-term children borne by a woman, excluding miscarriages or abortions, but including stillbirths.

Perinatal mortality: The frequency of stillbirths (fetuses of at least 28 weeks of gestation) plus deaths within the first week after birth.

Prevalence: The proportion of individuals in a population having a disease, exposure, or some other characteristic.

Rad: A unit of radiation absorbed dose equal to an energy deposition of 100 erg^3 per gram of material. 100 rad are equivalent to one Gray.

Radiochemical separation: Separation by chemical procedures of an element that contains a radionuclide of interest from a sample of a radioactive mixture.

Radionuclide: A radioactive, unstable form of a chemical element.

Reliability: The degree to which the results of a measurement procedure are stable and can be reproduced. Also related to validity (see below), in that an unreliable measurement cannot be valid.

Screening: Examination of people with no symptoms to detect unsuspected disease.

Screening effect: Typically, when a population is screened for a disease, more cases are found than would otherwise have come to medical attention, especially in the case of thyroid diseases. One study found that thyroid screening increased the number of cases of thyroid cancer found by about a factor of 8 and of benign thyroid nodules by about a factor of 15.

Sensitivity analysis: An analysis that describes the impact of changes in assumptions about uncertain or unknown aspects of a study on the design or results of the study.

***SMR* (*s*tandardized *m*ortality *r*atio):** The ratio of the number of deaths observed in a study population to the number of deaths expected if that population had death rates equivalent to those in some standard, general population (such as the US population). SMRs are typically calculated by using general population rates broken down by intervals of age and calendar time and by age and race.

Source term: A description of the amount of radionuclides or chemicals released from a site to the environment over a specific period that is used in dose reconstruction.

^{89}Sr: The isotope of strontium with a total number of protons and neutrons of 89.

Statistical power: The probability that a given study (such as the HTDS) will reject the null hypothesis of no effect of exposure on disease risk. Statistical power depends on the strength of the true association between risk and exposure, the number of participants in the study, and the distribution of exposure in the population being examined. Other issues peculiar to a given study (such as length of followup, degree of dose-measurement error, and subject compliance with study protocol) also affect statistical power.

Stratification: Division of a study population into groups (strata) for the purpose of performing a stratified analysis.

Stratified analysis: An analysis that estimates dose-response relationships separately for different groups (strata) of study participants and pools the results to form a single estimate.

Stratosphere: A relatively stable layer of the atmosphere between the tropopause and a height of about 30 miles in which the temperature changes little (in polar and temperate zones) or increases (in the tropics) with increasing altitudes. (See *Troposphere*.)

Subjective probability distribution function: A graph or formula that represents either an individual's own opinion or the consensus of expert opinion concerning the likely range of values of an unknown or uncertain quantity.

Thyroid-antibody test: A blood test that measures antibodies against the patient's thyroid tissue.

Thyroid burden: The total activity of a radionuclide in the thyroid.

Thyroid palpation: The procedure in which a physician characterizes the size, shape, and texture of the thyroid gland by manual examination of the neck.

Transfer coefficient to milk: The fraction of an element ingested daily by a cow that is secreted in milk when the intake of the element is at a steady state or equilibrium.

Troposphere: The region of the atmosphere between Earth's surface and the stratosphere in which the temperature falls with increasing altitude. The tropopause, the boundary between the troposphere and the stratosphere, normally occurs at an altitude of about 5-9 miles in polar or temperate zones and about 11 miles in the tropics.

Validation study: A special study designed to determine how valid an estimate or measurement is (see also Validity).

Validity: In general, the degree to which a measurement or an estimate measures what it purports to measure. There are several types of validity; the one most pertinent to this report is predictive validity, the accuracy of estimates in predicting actual measurements.

Information on Committee Members

ROY E. SHORE, Ph.D., Dr. P.H., (Chair) is a professor of Environmental Medicine and Director of the Epidemiology and Biostatistics Program at New York University School of Medicine. Dr. Shore received his Ph.D. degree from Syracuse University in 1967 and his Doctorate in Public Health from Columbia University in 1982. His research interests include environmental and occupational epidemiology, radiation epidemiology, and epidemiologic methods. He is on the standing committees on radiation biology/risk assessment of both the International Commission on Radiological Protection and the National Council on Radiation Protection and Measurements. He has served on several scientific advisory groups for the National Cancer Institute, the Department of Energy, and the Environmental Protection Agency, and on editorial advisory boards of the Journal of the National Cancer Institute, and Cancer Epidemiology, Biomarkers and Prevention.

BRUCE B. BOECKER, Ph.D., is a former Assistant Director of the Inhalation Toxicology Research Institute, Lovelace Biomedical and Environmental Research Institute, in Albuquerque, NM. He is currently a Scientist Emeritus at the Lovelace Respiratory Research Institute. Dr. Boecker earned his in Ph.D. in Radiation Biology from the University of Rochester and has conducted research at Lovelace since that time. His research interests lie

mainly in two broad areas, namely inhalation toxicology and dose-response relationships for long-term biological effects produced by internally deposited radionuclides. He has been particularly interested in the conduct of animal experimentation to develop information that may be used to predict the consequences of accidental exposure to humans and to establish standards that ensure the safe and orderly conduct of activities that may result in release of toxic agents to the environment. His personal research efforts have been associated primarily with the toxicology of airborne material associated with different activities in the nuclear fuel cycle. This research has spanned broadly from studies of aerosol characteristics as they may influence patterns of deposition, retention, and dosimetry through risk assessments for different nuclear energy systems. Dr. Boecker is also a Certified Health Physicist and has received a Distinguished Scientific Achievement Award from the Health Physics Society.

ANDRE' BOUVILLE, Ph.D., is a Senior Radiation Physicist in the Radiation Effects Branch of the National Cancer Institute. He earned the French equivalent of a Ph.D. at the University of Paul-Sabatier in Toulouse. Dr. Bouville's field of interest is radiation dosimetry and the environmental transfer of radionuclides. He has worked for the French Atomic Energy in several capacities including having been Assistant to the Director of Protection. Dr. Bouville was also Scientific Secretary for the United Nations Scientific Committee on the Effects of Atomic Radiation. Dr. Bouville is the member of several committees (including Committee 2 of the International Commission on Radiological Protection) and professional societies such as the National Council on Radiation Protection and Measurements.

A. BERTRAND BRILL, M.D., Ph.D., is a Research Professor in the Departments of Radiology and Physics at Vanderbilt University. Dr. Brill earned is M.D. at the University of Utah and his Ph.D. in Biophysics at the University of California, Berkeley. He served in the U.S. Public Health Service (PHS) in Japan at the Atomic Bomb Casualty Commission (ABCC) in the Statistics and Medicine Departments (1957–59), and as the PHS representative to ABCC until 1964. Dr. Brill's specialty is nuclear medicine and his major research areas include radiation leukemogenes, effects of radiation on thyroid function, and effects of diagnostic radioisotope studies, particularly exposures from I-131. Dr. Brill is currently a member of the NCI Task Group studying effects of the Chernobyl Accident on thyroid cancer induction in children. He was a former Medical Director, Division of Radiological Health, US Public Health Service, and a former Professor of Radiology, State University of New York at Stony Brook. He is a member of the Society of Nuclear Medicine Radiation Effects Committee, which he chaired for 10 years, the Medical Internal Radiation Dose Committee (MIRD), and the American Thyroid Association.

PATRICIA A.H. BUFFLER, Ph.D., is Dean Emerita, School of Public Health, University of California, Berkeley. Her current research interests in epidemiology include studies of leukemia in children, health effects of environmental tobacco smoke and health effects of non-ionizing radiation. She has served on numerous national and international advisory groups including advisory committees to the Department of Energy, the Department of Defense, the Department of Health and Human Services, the Environmental Agency, the Office of the President, the National Research Council and the World Health Organization. Since 1996 she has served as a Visiting Director for the US-Japan Radiation Effects Research Foundation. She has served as President for the Society of Epidemiologic Research, the American College of Epidemiology, and the International Society for Environmental Epidemiology and

is currently an officer of the Medical Sciences Council of the American Association for the Advancement of Science. She was awarded the American College of Epidemiology Lilienfeld Award in 1996. She is a Fellow of both the American College of Epidemiology and the Association for the Advancement of Science and a member of the Institute of Medicine/National Academy of Sciences.

SHARON M. FRIEDMAN, MA, is the Iacocca Professor and Director of the Science and Environmental Writing Program, at Lehigh University. She served as Chairperson of the Department of Journalism and Communication at Lehigh from 1986–1995. Her research focuses on how scientific, environmental, technological, and risk issues are communicated to the public. She served as a consultant to the President's Commission on the Accident at Three Mile Island and the United Nations Economic and Social Commission for Asia and the Pacific. She co-authored the book Reporting on the Environment: A Handbook for Journalists, which has been translated into 11 languages and widely distributed. She served as a Fulbright Distinguished Lecturer in Brazil and a Bosch Foundation Lecturer in Germany. Professor Friedman is the coeditor of the books, Communicating Uncertainty: Media Coverage of New and Controversial Science, and Scientists and Journalists: Reporting Science as News; Associate Editor of the journal Risk: Health, Safety & Environment. She is a Fellow of the American Association for the Advancement of Science (AAAS) and a member of the Council and Committee on Council Affairs of the AAAS. She is also chairperson of the Department of Energy's Low Dose Radiation Research Program Advisory Committee. Professor Friedman is a charter member of the Society of Environmental Journalists and a lifetime member of the National Association of Science Writers.

SUSAN E. LEDERER, Ph.D., is Assistant Professor in the School of Medicine, Section of the History of Medicine at Yale University. She received her doctorate in the history of science from the University of Wisconsin, Madison. A historian of American medicine, she served as a member of the President's Advisory Committee on Human Radiation Experiments. The author of Subjected to Science: Human Experimentation In America before the Second World War, she has written extensively on issues related to human and animal experimentation.

CARL M. MANSFIELD, M.D., is the chairman of the Radiation Oncology Department at the University of Maryland Medical Systems. His research interest has been in the treatment of cancer with emphasis on breast cancer. Dr. Mansfield has done extensive research in radiation dosimetry and brachytherapy. From 1976 to 1983, Dr. Mansfield was the chairman of the Department of Radiation Oncology at the University of Kansas. From 1983 through 1995, Dr. Mansfield was Professor and Chairman of the Department of Radiation Oncology and Nuclear Medicine at Thomas Jefferson University Hospital. From 1995 to 1997, he was the Associate Director of the Radiation Research Program at the National Cancer Institute. Dr. Mansfield is a Fellow of the American College of Radiology, the American College of Nuclear Medicine, and the Philadelphia College of Physicians. Dr. Mansfield has served on committees for the National Cancer Institute and the National Research Council.

DONALD E. MYERS, Ph.D., is Emeritus Professor of Mathematics at the University of Arizona and an adjunct Professor of Hydrology, adjunct Professor of Watershed Management, and a member of the faculty of the Applied Mathematics Program. He is a member of the University's Committee on Remote Sensing and Spatial Analysis and the Committee on Global Change. He earned his doctoral degree at the University of Illinois. His research has

included studies pertaining to the environmental restoration project at Los Alamos. He spent sabbaticals at the Centre de Geostatisque in Fontainbleau, France, and Stanford University. He held a visiting appointment at the Universite Paris XII and at the Centre de Geostatisque. He was a consultant to the National Research Council's BRER Committee on Exposure of the American People to I-131 from Nevada Atomic-Bomb Tests. Dr. Myers was an invited participant in the Project Varenius Workshop (NCGIA) and will be an invited participant in the NCEAS workshop on ecology and spatial analysis in the summer of 1999.

DANIEL O. STRAM, Ph.D., is an Associate Professor in the Department of Preventive Medicine at the University of Southern California. Dr. Stram earned his Ph.D. in Statistics from Temple University, and engaged in postdoctoral research in Biostatistics at the Harvard School of Public Health. From 1986–89 has was a member of the Statistics Department of the Radiation Effects Research Foundation in Japan. Since 1990, Dr. Stram's research interests have focused on clinical research and epidemiology in childhood and adult cancers at the University of Southern California and the Children's Cancer Group. His radiation-related work in Hiroshima and U.S.C. has concentrated on statistical aspects of dosimetry systems used for the A-bomb survivors and for the U.S. Uranium miner cohort study. Dr. Stram is a member of the Board on Radiation Effects Research (BRER) of the National Research Council.

ROBERT G. THOMAS, Ph.D., formerly of the Los Alamos National Laboratory, is a private consultant involved in lectures and workshops concerning the decommissioning, decontamination, and restoration of nuclear facilities. He attended the University of Rochester on a fellowship in radiological physics and subsequently received his Ph.D. in Radiobiology and Biophysics. Dr. Thomas was one of the planners and implementers in establishing the Inhalation Toxicology Research Institute in Albuquerque. He was

an Assistant Professor at the University of Rochester and an Adjunct Professor at the University of New Mexico. His research interests focused on establishing acceptable guidelines for exposure to radionuclides. He led a team of radiological health experts into Romania, Russia, and the Ukraine immediately following the Chernobyl accident. Dr. Thomas is currently on committees for the National Council for Radiation Protection and Measurements and for the International Commission on Radiological Protection.