

A Technical Review of the Final Report of the Hanford Thyroid Disease Study

March 30, 2004

A. James Rутtenber, Ph.D., M.D.  
F. Owen Hoffman, Ph.D.  
Raymond J. Carroll, Ph.D.  
Duncan C. Thomas, Ph.D.  
Sander Greenland, M.A., M.S., Dr.P.H., C. Stat

## Summary of Findings

We conclude that the results and conclusions of the Final Report of the Hanford Thyroid Disease Study (HTDS) (Davis et al., 2002) cannot be used to rule out important risks for thyroid cancer, neoplasms, or hypothyroidism from exposures to iodine-131 (I-131) from the Hanford nuclear facility. Specifically, we find that:

- (1) The interval estimates given by HTDS are much too narrow because they ignore large sources of uncertainty; in particular:
  - (a) They ignore major sources of uncertainty in dose assignment, and
  - (b) They do not account for the large losses from the cohort (the 1/3 for whom clinical outcome data were unavailable); and
- (2) The HTDS has misinterpreted its own statistics by relying on statistical significance, and then, on top of that, not calculating or interpreting significance test results correctly:
  - (a) They miscalculate significance levels by making inappropriate Bonferroni adjustments,
  - (b) They misinterpret nonsignificance as lack of evidence of a dose response, when in fact there is a trend in dose-response in some analyses, and
  - (c) They do not take adequate account of their own confidence intervals in formulating their closing statements.

Proper accounting for these problems would reveal that the study does not provide evidence capable of discriminating between no effect and relatively strong effects.

## Introduction

In this report, we address the extent to which the HTDS results and interpretations can be used to assess disease causation for individuals exposed to Iodine-131 (I-131) from the Hanford nuclear facility. We focus on three important questions: 1) Were radiation doses to the thyroid and their uncertainties modeled appropriately for the dose-response analyses that were conducted?; 2) Did the study, as conducted and analyzed, have adequate power to detect any likely effects at the chosen (0.05) significance level, as the HTDS authors assert?; and 3) Given the problems with estimates of doses and their uncertainties, and with the consequent low statistical power, were the analytic approaches and interpretations of results adequate and appropriate?

As detailed below, the answers to these questions are: "No," "No," and "No." For (1), we identify a number of problems with the dose estimates made for individual subjects, and substantial unmodeled sources of uncertainty in the radiation dose estimates. For (2), we point out that the statistical power for the study has been mischaracterized and is likely to be far lower than claimed by the HTDS. For (3), we identify problems with

the analytic choices that only further reduced the power, and show that the authors misinterpret lack of statistical significance as lack of evidence.

### **Hanford Dosimetry Uncertainty**

Ideally, dose reconstruction is based on measurements of the exposures received by individual members of the cohort or by measurements of the amount of radioactivity in air, deposition, and foodstuffs. Because of the absence of detailed environmental measurements and individual-specific exposure data, the HTDS relied almost exclusively on mathematical models to simulate the release, environmental transport, exposure and individual thyroid dose to each individual in their cohort. Such a practice is known to be associated with large and complex uncertainty (Hoffman 1991, Hoffman et al, 1993, Hoffman et al, 1996, Hoffman 1999).

When modeled doses are used as surrogates for measured values, it is important to characterize these uncertainties and account for them in the modeling process. Moreover, the full nature and extent of the uncertainty in individual dose estimation should be taken into account explicitly prior to determination of the statistical power of an epidemiological study. Upon the ascertainment of the incidence of disease among members of the cohort, the effect of dosimetry uncertainty in determining the slope of the dose response and the limits of its confidence intervals must also be considered.

The Hanford Environmental Dose Reconstruction Integrated Codes (HEDRIC) attempted to account for errors in dosimetry through a Monte Carlo approach. Probability distributions were assigned to uncertain model parameters, and 100 random alternative realizations of true dose were simulated for the entire cohort of 3191 "in area" study participants. Nevertheless, for individual dose assignment in HTDS some important sources of uncertainty were ignored. Some sources of uncertainty were inappropriately assigned default point estimates (single values that were held constant), some led to a systematic tendency to over and/or underestimate true dose for specific subgroups of the HTDS cohort, and some resulted in unrecognized complex mixtures of classical and shared and unshared Berkson errors. The incomplete handling of uncertainty in the HTDS has resulted in overestimates of statistical power and estimates of confidence intervals that are narrower than they should be.

The statistical theory of exposure measurement error distinguishes two basic ways such errors can come about. The "classical error" model assumes one is measuring exposure with some instrument (e.g., a dosimeter or questionnaire) which returns a value that is distributed in some fashion around the true (unknown) exposure with some error. The "Berkson error" model arises in experimental studies where an investigator aims to apply particular exposures to members of various groups but, due to inter-individual variability, the actual exposures received by individuals within each group is distributed in some fashion around the applied dose for the group.

In the standard theory, both kinds of errors are assumed to have zero mean and to follow some known distribution (e.g., normal). The standard theory also assumes that measurement errors are uncorrelated between individuals. A consequence of classical measurement error is that the slope of an exposure-response relationship is attenuated — biased towards the null — by a factor that depends upon the relative variance of true exposures and measurement errors; in contrast, for certain types of models, Berkson error may produce no bias towards the null, but the variance of the slope estimate will be increased (and power reduced) by a factor similarly depending upon the relative variance of applied exposures and within-group variability.

In epidemiologic studies such as the HTDS, the measurement system is complex and contains elements of both kinds of error. Calculated doses (conditional on such variables as date of birth, place of residence, and dietary habits) are similar to the “applied exposures” in the experimental setting and are interpreted as expressing the mean (or median) of a distribution of unknown true doses for all individuals with similar characteristics; thus, they can be expected to have a Berkson error structure. On the other hand, the questionnaire information that forms one of the inputs to the dosimetry calculations is more likely to have a classical error structure.

To further complicate the situation, measurement errors are likely to be highly correlated between individuals to the extent that they share certain characteristics — e.g., the errors for two people living in the same town will be similar if the assumed deposition in the town was too high or too low. The quantitative implications of mixed Berkson and classical errors and of shared errors will be discussed below.

The full nature and extent of the uncertainty in individual dose estimation must be taken into account in order to correctly determine the statistical power of the analyses presented in the HTDS. The effect of dosimetry uncertainty in determining the slopes for the dose-response models and the limits of its confidence intervals must also be considered before these estimates can be appropriately interpreted.

#### Inter-individual Stochastic Variability

Inter-individual variability of true dose within the HTDS cohort can be substantial due to the heterogeneity of I-131 deposition, interception by vegetation, food chain transport, individual variations in consumption rates, and person-to-person variations in thyroid mass and fractional uptake from blood to the thyroid gland. It appears from the documentation of HTDS and from correspondence with Bruce Napier that the only model variables that explicitly accounted for inter-individual variability of true dose, given commonalities of age, gender, location, and dietary source of the individual, were the uncertainties assigned to the dose conversion factor and the GSD assigned to the individual’s intake of foods.

Within the HEDRIC/STRM/RATCHET/DESCARTES/CIDER suite of codes used for the HTDS dose reconstruction, it appears as if the dose conversion factor (DCF) and the default dietary consumption rates were the only uncertain variables that are allowed to vary independently from individual to individual within a given realization. All other parameter uncertainties in the HEDR suite of codes were strongly correlated and thus shared among individuals within the cohort (per each of the 100 realizations of the cohort dose). Therefore, within any realization of the dose for all members of the HTDS cohort, the true inter-individual variability of dose in the HTDS cohort will be substantially understated and the degree of correlation of uncertainty among individuals will be overstated, i.e., the rank order of dose assignments among individuals will remain nearly the same from realization to realization for individuals consuming the given milk type or source. Thus, the mean cohort dose will vary substantially from realization to realization, but the relative inter-individual variability of dose will not. This effect could lead to an overestimation of the uncertainty in the cohort mean dose (which is not reported in HTDS) and an underestimation of the relative variance of the cohort dose.

This problem could have been remedied by using a two dimensional Monte Carlo procedure to separate uncertainty due to stochastic inter-individual variability from uncertainty due to lack of knowledge about true fixed quantities and true geometric means (GMs) and geometric standard deviations (GSDs) that define stochastic variation of true dose between members of the HTDS cohort [see IAEA (1989) and NCRP Commentary No. 14 (1996) for details].

For each of the 100 alternative realizations of the true distribution of cohort doses, unique GMs and GSDs would be specified from a random sample from probability density functions (PDFs) that represent uncertainty in these quantities. In other words, the GMs and GSDs that define stochastic inter-individual variability of true dose themselves are uncertain. Therefore, in each realization, a unique value of GM and GSD would be sampled from PDFs representing uncertainty in the true values of GM and GSD. From the unique GM and GSD thus obtained, an estimate of true stochastic variability of within-cohort doses would be simulated, so that the cohort mean dose and the relative variability of individual doses would vary from realization to realization. Within a realization, the inter-individual cohort GSD would thus be increased from what is currently calculated within HEDR.

#### Bias and Uncertainty for Milk Pathway Dosimetry

*Commercial Cow-Milk Transfer Coefficient:* It is well documented that the distribution assumed in HEDR for commercial cow milk (0.012 d/L) is biased on the high side from what has been reported in the literature and includes an uncertainty that is unrealistically small (standard deviation, 0.002 d/L) (Snyder et al. 1994; Napier et al,

2000; Nuclear Regulatory Commission, 1997). This issue was raised in the 1993 peer review conducted for Pacific Northwest Laboratory (Hoffman et al., 1993) and by the NAS review of HTDS (NRC, 2000). However, the authors of HEDR (Napier et al, 2000) chose not to revise their assumptions and declined to examine the effect of changes in these assumptions on the power of the study. Had the assigned distributions for the commercial milk transfer coefficient been wider and centered on values that were more consistent with central estimates reported in the literature, the true dose would be lower and the uncertainty higher for those consuming commercial sources of fresh milk. This source of bias in the dose estimate has not been accounted for within HTDS. The overall effect of this source of bias is to overstate the statistical power of the study and understate the slope and confidence interval of the dose response.

Note that in the I-131 fallout dose reconstructions of the University of Utah and the National Cancer Institute (Stevens et al., 1992, NCI 1997), the commercial milk transfer coefficient for I-131 was assumed to be about 0.004 d/L with a GSD approaching 2.0. The expert elicitation performed by the Nuclear Regulatory Commission (1997) show that uncertainty on the average milk transfer coefficient for a very large region (subsequent to an accidental release of elemental I-131) could be as low as 0.0005 d/L and as high as 0.05 d/L, with most experts reporting uncertainties of the mean value extending down to 0.001 d/L.

The standard deviation of 0.002 d/L assumed in HEDR for the milk transfer coefficient of commercial cows leads to a coefficient of variation of merely 17%, which translates to a GSD of about 1.18. This is a much smaller uncertainty than what would be indicated by GSDs typically assumed for this parameter that range from about 1.4 to 2.1 (Stevens et al., 1992; NCI, 1997; Apostoaei et al., 1999, 2003).

In responding to the NAS review of the draft report of HTDS, the authors of HEDR (Napier et al. 2000 [Appendix 22 of HTDS]), appear to have confused a frequency distribution that describes stochastic variability of true values of the milk transfer coefficient with a distribution that is intended to represent uncertainty for an average value that would prevail throughout a large region of the USA and Europe. Although they acknowledged that their original distribution did not include the lower values reported in the expert elicitation, they chose not to revise their assumptions on the basis that such a revision would not substantially affect the mean value resulting from the use of their original assumptions for HTDS. This logic results in a bias towards overestimation of mean dose to the HTDS subcohort that consumed commercial sources of fresh milk and an underestimation of uncertainty in the doses assigned to these persons.

A systematic overestimation of true dose and a systematic underestimation of the dose uncertainty for this subgroup could compromise the reported HTDS estimates of

statistical power. These errors in estimation could also reduce the central value and the upper confidence limit of the slope of the HTDS reported dose response. Furthermore, the resulting systematic misclassification of dose for specific subgroups could be differentially distributed, thereby obscuring the relationship between dose and risk in dose-response models.

*Mother's Milk Transfer Coefficient:* There is evidence that the milk transfer coefficient assumed in HEDR for the transfer of I-131 into mothers' milk has been underestimated. This underestimation would lead to an underestimate of true dose for those who were breast-fed as infants, especially during 1945 (the year with the highest releases from Hanford). This systematic underestimation of true dose would lead to differential misclassification of disease outcomes for a subgroup of the HTDS cohort, which in turn would impact the power of the study and the evaluation of the dose response.

The milk transfer coefficient for mother's milk used in the HEDR suite of models was assumed to be uniform from 0.07 to 0.36 d/L, based on limited data in the literature (Snyder et al., 1994). More recent information from Simon et al., (2002), based on a more extensive review of measurements, gives a lognormal distribution with a GM of 0.37 d/L and a GSD of 1.5. This updated distribution is used in the NCI individual dose calculator for nationwide exposures to NTS fallout I-131. The NTS individual dose calculator from NCI can be found at (<http://ntsi131.nci.nih.gov/>). It is also being applied within the University of Utah revised dose reconstruction model that is being finalized for the refined epidemiological analysis of the special cohort who lived in counties immediately downwind of the NTS.

The median value for the HEDR models for the mother's milk transfer coefficient is 0.22 d/L. This is a factor 1.7 less than the geometric mean value reported by Simon et al. (2002). The maximum value of 0.36 d/L assumed in HEDR is slightly less than the geometric mean value reported by Simon et al., and nearly a factor of 2.3 less than the upper 97% percentile of the lognormal distribution reported by Simon et al. In HTDS, the only individuals explicitly considered to have been on a diet of mother's milk were those for whom such information was provided in the computer-assisted telephone interviews (CATIs). These subjects would have had true doses that were underestimated. The extent of underestimation would depend on the magnitude of dose received in later years when the individual's diet changed to other sources of fresh milk.

*Default Assumptions for Subjects without CATIs:* There is the potential for a systematic bias towards overestimation of true dose for the 1212 participants in HTDS who were assigned a default diet of cow's milk in the absence of data from a CATI. The degree of overestimation of true dose could be considerable for the HTDS subgroup of subjects who were born in 1945 and who may not have consumed fresh sources of milk, but

were assigned a default diet of fresh cow's milk because of the lack of specific dietary information obtained from a CATI.

It appears, from the information from CATIs, that about 36% of the HTDS study subjects were not on a diet of fresh cow's milk in 1945, the year of the highest I-131 releases from Hanford. Overall, about 69% of the cohort was not on a diet of fresh milk before the age of 6 months, which dropped to about 15% at age 1 year.

As noted above, systematic overestimation of dose for a subgroup in HTDS would be expected to lead to an overestimation of the true statistical power of the study, and to have obscured possible dose-response relationships.

#### Underestimated Dose Uncertainty

*Residence Histories:* Uncertainties in HTDS dose assignments have been substantially underestimated by using default dates as surrogates for true dates of changes of residence history, changes in both diet and food sources. This procedure essentially treats the default dates as if they are known without error. No additional analysis of this source of uncertainty is included in HTDS.

Given that the dates of change of residence history and dates of change of dietary sources and amounts are inherently uncertain when interviews are conducted approximately 50 years since the time of initial exposure to Hanford I-131, probability distributions of possibly true dates should have been defined and this source of uncertainty included within the alternative realizations of the true doses

*Shared Uncertainty:* Uncertainties of the HTDS doses have also been underestimated by allowing discrete time periods and exposure pathways to be treated as uncorrelated when summing uncertain doses for a single person. This procedure is inappropriate as these time periods and exposure pathways have shared sources of uncertainty for estimation of the true dose to a specific person. This is because the estimated dose for an individual for any specific time period is composed of a fixed true component and random error, and the latter errors are likely to be positively correlated across periods within individuals. Therefore, the sum of the error variances from the periods underestimates the total error variance since it ignores the positive covariance terms.

For example, specific alterations to the original CIDER code was made to allow the uncertainty in the I-131 dose conversion factor to be random and independent per individual. The uncertainty in the dose conversion factor is also considered to be independent per time period and exposure pathway (i.e. ingestion versus inhalation) for a given person. Since the uncertainty in the I-131 dose conversion factor is dominated by the uncertainty in the mass of the thyroid gland, the assumption of

