

## THE HANFORD THYROID DISEASE STUDY: AN ALTERNATIVE VIEW OF THE FINDINGS

F. Owen Hoffman,\* A. James Rittenber,<sup>†</sup> A. Iulian Apostoaei,\* Raymond J. Carroll,<sup>‡</sup> and Sander Greenland<sup>§</sup>

---

**Abstract**—The Hanford Thyroid Disease Study (HTDS) is one of the largest and most complex epidemiologic studies of the relation between environmental exposures to <sup>131</sup>I and thyroid disease. The study detected no dose-response relation using a 0.05 level for statistical significance. The results for thyroid cancer appear inconsistent with those from other studies of populations with similar exposures, and either reflect inadequate statistical power, bias, or unique relations between exposure and disease risk. In this paper, we explore these possibilities, and present evidence that the HTDS statistical power was inadequate due to complex uncertainties associated with the mathematical models and assumptions used to reconstruct individual doses. We conclude that, at the very least, the confidence intervals reported by the HTDS for thyroid cancer and other thyroid diseases are too narrow because they fail to reflect key uncertainties in the measurement-error structure. We recommend that the HTDS results be interpreted as inconclusive rather than as evidence for little or no disease risk from Hanford exposures.

Health Phys. 92(2):99–111; 2007

Key words: thyroid; epidemiology; <sup>131</sup>I; health effects

---

### INTRODUCTION

THE HANFORD Thyroid Disease Study (HTDS) (Davis et al. 2002, 2004) examined a large number of dose-response relationships for thyroid disease endpoints using multiple methods of analysis. The study detected no dose-response relation using a 0.05 level for statistical significance. The authors concluded that the cancer risk for children exposed to <sup>131</sup>I from Hanford releases is much lower than from similar doses from external exposures to penetrating radiation, from internal exposures to shorter-lived isotopes of iodine, or from exposures to <sup>131</sup>I after the Chernobyl accident. In companion

---

\* SENES Oak Ridge, Inc., Center for Risk Analysis, Oak Ridge, TN 37830; <sup>†</sup> University of Colorado School of Medicine, Denver, CO; <sup>‡</sup> Texas A & M University Department of Statistics, College Station, TX; <sup>§</sup> Departments of Epidemiology and Statistics, University of California, Los Angeles.

For correspondence contact: F. Owen Hoffman, SENES Oak Ridge, Inc., Center for Risk Analysis, 102 Donner Drive, Oak Ridge, TN 37830, or email at [sensor@senes.com](mailto:sensor@senes.com).

(Manuscript accepted 5 July 2006)  
0017-9078/07/0

Copyright © 2007 Health Physics Society

papers, the authors maintained they had adequate statistical power to exclude false-negative results (Stram and Kopecky 2003; Davis et al. 2004; Kopecky et al. 2004, 2005).

As discussed below, these findings are inconsistent with those from studies of other populations with environmental exposures to <sup>131</sup>I and raise important issues for radiation health research. Moreover, it is difficult to reconcile the HTDS results with the evidence for thyroid disease that has been reported for the Chernobyl accident, which also includes exposures primarily to <sup>131</sup>I.

With this paper we hope to generate a dialogue on these issues. To begin the discussion, we describe the inconsistencies between the findings of the HTDS and other studies and the explanations that have been proposed for these differences. We then propose steps that can be taken to further clarify the issue, and suggest how the HTDS findings might be regarded by the scientific community in the interim.

### OUTLINE OF POSITIONS

The authors of the HTDS attribute the differences in excess risk estimation between HTDS and other studies to differences in the dose and dose rates delivered by Hanford <sup>131</sup>I, the absence of releases of shorter-lived radionuclides that would elevate the dose rate of the total dose incurred, and the fact that many children exposed downwind of Chernobyl were deficient in stable dietary iodine. They have maintained in companion papers that their study had adequate statistical power to detect an excess risk of disease that would be consistent with results from other epidemiological investigations of children exposed to external radiation, weapons fallout, and Chernobyl <sup>131</sup>I (Stram and Kopecky 2003; Davis et al. 2002, 2004; Kopecky et al. 2004, 2005). They contend that if an excess risk was present in the Hanford cohort, it was very low.

We believe that the anomalous findings of the HTDS can be explained by uncertainties associated with the

mathematical models and assumptions used to reconstruct individual doses, which were more extensive than acknowledged. In their analysis of statistical power, the authors of the HTDS made the assumption that all uncertainties are non-differential, unshared Berkson measurement errors. Correcting for the presence of differential, systematic over-estimation of individual dose with respect to the model covariates that determine individual dose would result in a reduction of the statistical power of the HTDS and an increase in the central value of the dose-response relationship, as well as an increase in the upper bound of the interval estimate. Accounting for the presence of large non-differential uncertainties (composed of mixtures of shared and unshared Berkson and classical measurement errors) results in a further reduction of the statistical power, an increase in the central values of the dose-response relationship, and a further widening of interval estimates about the excess relative risk (ERR) per unit dose (Carroll et al. 2006).

Correcting for the presence of systematic and random sources of uncertainty would cause the intervals to overlap central estimates of the dose-response relationship reported for other epidemiologic studies of radiation-induced thyroid cancer. This overlap is particularly evident when the original dose-response data from exposure to external sources of radiation are adjusted for the effects of low doses and low dose rates through division by a distribution of values reported for the dose and dose-rate effectiveness factor (DDREF).

We believe that available scientific evidence does not support the assumption that the very large differences in dose response between the HTDS and other studies were caused by differences in the rates and types of exposure, the mixture of radionuclides involved, or the amount of stable iodine in the diet.

### CONFLICTING EVIDENCE

Although the carcinogenicity of  $^{131}\text{I}$  has been debated for decades, the issue has been settled by recent data from the Chernobyl accident. Studies of children exposed to  $^{131}\text{I}$  from Chernobyl indicate that the risk for thyroid cancer increases with dose and that the quantitative estimates of excess risk obtained to date are not substantially different from excess risk estimates obtained from cohorts exposed to external sources of radiation (IOM/NRC 1999; UNSCEAR 2000) (Table 1). Although iodine deficiency appears to increase the ERR for  $^{131}\text{I}$ -induced thyroid cancer by about a factor of three (Cardis et al. 2005), the risk from  $^{131}\text{I}$  is independent of that from iodine deficiency (Shakhtarin et al. 2003;

Cardis et al. 2005; Tronko et al. 2006). For the populations exposed to Chernobyl radiation releases, the contributions to thyroid dose from mixtures of both short-lived radionuclides (e.g.,  $^{132}\text{I}$ , radiotelluriums) or long-lived radionuclides (e.g.,  $^{137}\text{Cs}$ ) were less than a few percent (Gavrilin et al. 2004; Cardis et al. 2005; Minenko et al. 2006).

There is also evidence supporting a causal relation between  $^{131}\text{I}$  and thyroid neoplasia from follow-up studies of children exposed to fallout from atmospheric testing conducted at the Nevada Test Site (NTS) (Table 1; Mallick et al. 2002; Lyon et al. 2006; Simon et al. 2006). The estimates of excess risk from this population are similar to those identified for children exposed to external gamma and x rays from medical procedures; the similarity becomes stronger when the excess risks observed from acute exposure to high energy gamma rays and fractionated x rays are adjusted for low-dose and low dose-rate effects (Table 1).

In their recent report, the BEIR VII Committee (NRC/NAS 2006) concluded that the most realistic model for estimating radiogenic cancer risk is linear, with an ERR per unit dose that decreases with age at exposure. For thyroid cancer, the BEIR VII preferred model is:

$$\text{ERR per Gy} = 0.53 \exp[-0.083(e - 30)] \text{ for males,} \quad (1)$$

and

$$\text{ERR per Gy} = 1.05 \exp[-0.083(e - 30)] \text{ for females,} \quad (2)$$

where  $e$  is the age at time of exposure. For chronic exposures to low linear energy transfer (LET) radiation such as  $^{131}\text{I}$ , BEIR VII recommended dividing the ERR in the above equations by a DDREF that is described by a lognormal distribution with a geometric mean (GM) of 1.5 and an adjusted geometric standard deviation (GSD) of 1.35. The overall uncertainty in the ERR per Gy/DDREF is approximated by a GSD of about 2.0.

For thyroid cancer, the risk coefficients developed by BEIR VII were derived from epidemiologic studies of children and adolescents exposed to external sources of penetrating radiation (Ron et al. 1995). The BEIR VII committee relied on data from external exposures because risk coefficients based on epidemiologic data from populations exposed to internal sources of radiation are highly uncertain.

The BEIR VII committee concluded: "Although there are no strong reasons to think that the dose-response from internal low-LET exposure would differ

**Table 1.** Results of epidemiologic studies of thyroid cancer in children exposed to fallout <sup>131</sup>I or external radiation.

Study (by exposure type)	ERR/Gy 95% confidence interval			Average dose (Gy)	Age at exposure (y)	Reference
	2.5 <sup>th</sup> percentile	50 <sup>th</sup> percentile	97.5 <sup>th</sup> percentile			
<b><sup>131</sup>I exposure studies</b>						
Hanford Thyroid Disease Study						
Hanford region	0.0	0.48	4.6	0.174	0–5	Davis et al. (2002)
Nuclear weapons fallout						
Across entire U.S.	2.8	12	31	0.05 <sup>a</sup>	1955 birth cohort	Gilbert et al. (1998) <sup>b</sup>
Across entire U.S.	2.7	13	36	0.05 <sup>a</sup>	1955 birth cohort	Gilbert et al. (1998) <sup>c</sup>
Across entire U.S.	–1.1	10.6	29	0.043	0–1	Gilbert et al. (1998) <sup>b</sup>
Across entire U.S.	–0.5	2.4	5.6	0.043	0–1	Gilbert et al. (1998) <sup>d</sup>
Utah, Nevada, and Arizona	2.7	13.02	68.7	0.12	2 (average)	Lyon et al. (2006) <sup>e</sup>
<b>Chernobyl</b>						
Bryansk, Kaluga, Orel and Tula Regions, Russia	2.4	7.3	12	0.056	0–17 (females)	Ivanov et al. (1999)
Bryansk Region, Russia	0.1	1.7	3.2	0.69—females <sup>f</sup> 0.48—males <sup>f</sup>	0–19	Davis et al. (2004)
Bryansk Region, Russia	0.1	2.7	5.3	0.555 <sup>f</sup>	0–2 in 1986	Davis et al. (2004)
Ukraine and Belarus	8	10	12	Not reported	0–18	Jacob et al. (2004)
Belarus and Bryansk Region, Russia	8.6	23	82	0.049 Bryansk, city 0.37 Gomel, city 0.07 Minsk, city	0–19	Jacob et al. (1999)
Belarus and Russia	1.2	3.9	6.5	0.365 Belarus <sup>f,g</sup> 0.04 Russia <sup>f,g</sup>	0–15	Cardis et al. (2005)
Belarus	1.9	4.1	7.9	0.356 <sup>f,g</sup>	0–15	Cardis et al. (2005)
Russia	0.3	30.5	760	0.04 <sup>f,g</sup>	0–15	Cardis et al. (2005)
Bryansk Region, Russia	4.8	48.7	1,151	0.0435 (cases)	0–19	Kopecky et al. (2006) <sup>h</sup>
Chernihiv, Zhytomyr, and Kyiv Regions, Ukraine	–0.36	138	5.4 × 10 <sup>8</sup>	0.016 (controls)		
	1.7	5.25	27.5	2.0 ± 2.52 (case) 0.78 ± 1.85 (non-case)	0–18	Tronko et al. (2006) <sup>i</sup>
<b>Gamma and x-ray exposure studies; adjusted for DDREF<sup>h</sup></b>						
Individual studies						
Japanese A-bomb survivors	2.0	6.5	17	0.3	0–10	Thompson et al. (1994)
X-ray therapy for enlarged thymus (Rochester, NY)	1.4	6.3	25	1.4	0–1	Ron et al. (1995)
X-ray therapy for tinea capitis (Israel)	6.3	22	61	0.09	0–15	Lubin et al. (2004)
X-ray therapy for tinea capitis (New York)	0.7	6.1	53	0.1	0–18	Shore and Xue (1999)
X-ray therapy for enlarged tonsils and adenoids (Chicago)	0.1	1.7	20	0.6	4 (average)	Ron et al. (1995)
X-ray therapy for lymphoid hyperplasia	1.2	4.1	13	0.2	0–15	Shore (1992)
Radiotherapy with beta, gamma and x rays for skin hemangioma (Stockholm cohort) <sup>j</sup>	1.3	4.9	10	0.1	infants	Lundell et al. (1994)
Radiotherapy with <sup>226</sup> Ra for skin hemangioma (Gothenburg cohort) <sup>j</sup>	0.4	7.5	18	0.1	infants	Lindberg et al. (1995)
Pooled analyses						
Pooled analysis of x-ray treated patients and A-bomb survivors	1.0	5.7	29		2	Land et al. (2003)
BEIR VII (female)	1.5	7.0	34		2	NRC/NAS (2006)
BEIR VII (male)	0.8	3.5	17		2	NRC/NAS (2006)

<sup>a</sup> Estimated based on the information presented by Gilbert et al. (1998).<sup>b</sup> Mortality: county-specific doses.<sup>c</sup> Mortality: state-specific doses.<sup>d</sup> Incidence: county-specific doses.<sup>e</sup> All uncertainties treated as Berkson, without systematic errors (see Lyon et al. 2006).<sup>f</sup> Median dose.<sup>g</sup> Dose from all radiation types. Doses are dominated by exposure to <sup>131</sup>I. Doses from <sup>131</sup>I are 0.356 Gy and 0.039 Gy, in Belarus and Russia, respectively.<sup>h</sup> ERR/Gy in the first row were obtained by ignoring the uncertainty in thyroid doses. ERR/Gy in the second row were obtained by treating all uncertainties in doses as classical measurement error.<sup>i</sup> All uncertainties in thyroid doses were treated as Berkson, without systematic errors.<sup>j</sup> A discrete distribution of DDREF was applied to ERR/Gy for those studies not involving radiotherapy for skin hemangioma where exposure was chronic. (DDREF values are 0.5, 0.7, 1.0, 1.5, 2.0, 3.0, and 4.0, respective probabilities for these values are 0.01, 0.04, 0.35, 0.23, 0.23, 0.1, and 0.04; Source: Land et al. 2003.)

from that for external exposure, there is additional uncertainty in applying the BEIR VII risk models to estimate risks from internal exposure" (NRC/NAS 2006). As noted above, when quantifying the risk of thyroid cancer from prolonged low-dose exposures to low-LET radiation, the BEIR VII committee adjusted its preferred risk model through division by a DDREF, which introduced both a bias correction and additional uncertainty associated with chronic exposure to low-LET radiation.

In light of the preceding evidence for thyroid cancer risk following  $^{131}\text{I}$  exposure, we conclude that the HTDS findings are either falsely negative or truly negative and explained by unique conditions. We explore each possibility in the following sections and suggest ways to clarify the controversy.

### DOSE MODELS AND STATISTICAL POWER

Based on analyses of statistical power, the HTDS authors maintain that their findings are not falsely negative (Davis et al. 2002, 2004; Kopecky et al. 2004, 2005). However, if these analyses did not account for all factors that could reduce power, then the HTDS assertions overstate the certainty of the results and the power of the HTDS might be substantially lower than the authors claimed.

In the HTDS, cumulative doses to the thyroid were estimated for the 3,191 subjects by using a series of mathematical models to calculate the release, downwind dispersion, deposition, food-chain contamination, inhalation and ingestion of  $^{131}\text{I}$ , and the absorption of energy resulting from the decay of  $^{131}\text{I}$  in each subject's thyroid gland (Farris et al. 1994; Davis et al. 2002; Kopecky et al. 2004). These models were linked with data obtained from individual exposure histories and estimates of the commercial distribution of fresh dairy products within the spatial and temporal domains of the HTDS. The information obtained for individual subjects in interviews and the assumptions made in characterizing the distribution of fresh dairy products are uncertain for a variety of reasons. These uncertainties, as well as those associated with other variables, such as estimates of  $^{131}\text{I}$  release, dispersion, deposition, and food-chain transport, can widen the interval estimates of dose and ERR, and lower the statistical power of a study.

The exclusive reliance upon models for reconstructing  $^{131}\text{I}$  exposures from the Hanford facility differs from the dose reconstructions for the NTS and Chernobyl. For the NTS study (Simon et al. 2006), modeled doses were based on external exposure measurements or on the quantification of radionuclides deposited on gummed film collectors. The mathematical models used to estimate the interception of  $^{131}\text{I}$  in NTS fallout by vegetation

and the subsequent  $^{131}\text{I}$  contamination of milk were somewhat similar to those used at Hanford, but Hanford modelers chose values for transfer coefficients that were markedly larger than those for the NTS, resulting in higher estimates of dose per unit deposition.

In contrast to the exclusive reliance upon mathematical models and assumptions about model coefficients by the HTDS, a variety of models and measurements were used to estimate doses for subjects in Chernobyl studies. These measurements included deposition of multiple radionuclides in soil and vegetation, radionuclide concentrations in milk and other foods, and measurements of radionuclide burdens in the thyroid glands of a large number of subjects. For the investigation of Chernobyl cohorts, models calibrated against available data were used to estimate individual exposures in those situations where direct measurements were unavailable (Gavrillin et al. 2004; Cardis et al. 2005; Tronko et al. 2006).

Below, we describe the sources of dose uncertainty in the HTDS in more detail, and discuss the extent to which they have been considered in analyses of statistical power for the HTDS.

### UNCERTAINTIES IN RESIDENCE HISTORIES AND DIETS OF SUBJECTS

Because HTDS subjects were infants or young children at the time of the largest  $^{131}\text{I}$  releases, the most reliable residence and dietary information was obtained from parents or other knowledgeable respondents. For 1,979 of the subjects (62%), detailed residence and dietary histories were obtained through computer-assisted telephone interviews (CATI) with these respondents (Kopecky et al. 2004). In the CATI interviews for those subjects who were born or breast-fed after the releases of  $^{131}\text{I}$  began, interviewers obtained a limited dietary history for the subjects' mothers to estimate doses from prenatal exposures and from ingestion of human breast milk.

For the 1,212 subjects (38%) who could not identify knowledgeable sources of historical information, an abbreviated in-person interview was administered to obtain residence history and to determine whether subjects consumed fresh milk or dairy products. These subjects were assigned the same rates of food intake, based on typical diets. Because the dietary histories and breast-feeding practices of the mothers of these subjects could not be obtained, it was assumed that cow's milk consumption began at birth. This assumption was not, however, supported by the HTDS data obtained from those subjects who were able to provide interviews, since 69% of the subjects with CATI interviews reportedly did not consume fresh dairy products in the first 6 mo of life

(Kopecky et al. 2004). The dose estimates for 38% of the study subjects were, therefore, far more uncertain than those for the remainder of the study population. This uncertainty includes the potential for bias in the dose estimate for these study subjects.

For the HTDS, the absence of reliable information on residence history, dietary intake of  $^{131}\text{I}$ -contaminated food, and the impact of breast-feeding on exposure to  $^{131}\text{I}$  all lead to increased uncertainty and the potential for both non-differential and differential misclassification of dose estimates for individual subjects and for groups of subjects. Individual-dose error that is independent of the true exposure, outcome, covariates, and other errors (classical error) would, under these circumstances, bias estimates of risk towards the null and obscure positive dose-response relations (Rothman and Greenland 1998; Carroll et al. 2006). In dose-response models, classical error and systematic overestimation of dose have similar effects—both induce bias towards underestimating the true dose-response relationship (NRC/NAS 2006).

The HTDS reported no evidence of bias associated with the lack of exposure information for the subjects without CATI interviews, and did not detect confounding by or heterogeneity across the sources of dosimetry data (Davis et al. 2002, 2004). However, the study authors did not consider the impacts of systematic overestimation of doses for the 38% of subjects without CATI interviews on (a) the modeling of dose-response relationships, (b) their confidence intervals (Davis et al. 2002, 2004), or (c) the estimates of statistical power for these models (Stram and Kopecky 2003; Kopecky et al. 2004).

The methods used for reconstructing individual exposure histories for Hanford subjects are different from those applied to other populations with environmental exposures to  $^{131}\text{I}$ . For the studies of children exposed to NTS fallout, residence and dietary histories were obtained directly from the parents of child subjects or from other knowledgeable sources within 15 y of exposure (Simon et al. 2006). In the Chernobyl studies with dose estimates for individual subjects, data for exposure reconstruction were also obtained through interviews, which were conducted within no more than a few years of exposure.

## UNCERTAINTIES IN DOSE MODELS

Dose estimates made with mathematical exposure models are known to be associated with large and complex uncertainties (Hoffman 1999; Mallick et al. 2002; Schafer and Gilbert 2006). These uncertainties include shared and unshared mixtures of Berkson (i.e.,

true doses independently distributed about assigned values) and classical measurement errors (i.e., dose estimates or measurements independently distributed about true values), as well as sources of bias leading to differential errors (errors that depend on covariates or outcomes) and dependent errors across subjects or groups of subjects.

For example, one of the most important parameters in dose reconstruction models for  $^{131}\text{I}$  is the coefficient used to represent the transfer from animal intake to milk ( $F_m$ ). The HTDS assigned median values to  $F_m$  of 0.012  $\text{d L}^{-1}$  for milk produced by commercial dairy cows, and 0.009  $\text{d L}^{-1}$  for backyard cows (Snyder et al. 1994). For the NTS study, the median value of  $F_m$  was about 0.004  $\text{d L}^{-1}$  for both sources of cow's milk (Stevens et al. 1992; NCI 1997). Recent model testing using Hanford data from 1963 indicates that the median value for  $F_m$  could be as low as 0.002  $\text{d L}^{-1}$  (Apostoaie 2005).

There is also an indication that the uncertainty in  $F_m$  was substantially underestimated in the HTDS. For commercial sources of milk, the uncertainty for  $F_m$  was constrained to a standard deviation of 0.0002  $\text{d L}^{-1}$  about a mean value of 0.012  $\text{d L}^{-1}$  (approximating a GSD of 1.18) (Snyder et al. 1994). For the NTS study, a GSD of close to 2.0 was assigned (Stevens et al. 1992). In their analyses of statistical power, the HTDS did not consider the impact of different, but possibly true values of  $F_m$  among various sources of commercial and non-commercial milk. Because thyroid dose is a linear function of  $F_m$ , it is likely that the HTDS thyroid doses were overestimated and their uncertainties underestimated for individuals who consumed commercial sources of fresh cow's milk. Problems associated with the choice of a distribution of values for  $F_m$  in the HTDS have been identified in previous reviews (Hoffman 1993; NRC/NAS 2000).

In this paper, we do not have sufficient space to discuss other sources of uncertainty and potential bias in the dose reconstruction data used in the HTDS. We have, however, summarized the important ones in Table 2.

## DOSE UNCERTAINTY AND VARIABILITY

In the HTDS, the uncertainty in dose estimation was represented by 100 unique sets, or vectors, of alternate dose realizations for each of 3,191 subjects. These realizations of dose were implemented with Monte Carlo sampling of the probability distributions assigned to the model inputs and coefficients for which uncertainties had been characterized. These 100 dose realizations approximated a lognormal distribution, characterized by a GM and a GSD. The average of the GSDs for uncertainty in

**Table 2.** Sources of uncertainty associated with the evaluation of HTDS dose estimates.

	Potential for bias/uncertainty
<b>Sources of bias</b>	
Central value for $^{131}\text{I}$ transfer coefficient for commercial milk	+++ <sup>a</sup>
Central value of $^{131}\text{I}$ transfer to mother's milk	— <sup>b</sup>
Percentage of individuals assumed to be on fresh milk diet in 1945 in the absence of a CATI interview	++
Amount of $^{131}\text{I}$ released after 1950	—
Confounding by Nevada Test Site $^{131}\text{I}$ dose	---
Misclassification of thyroid doses from Nevada Test Site $^{131}\text{I}$ <sup>c</sup>	---
<b>Sources of uncertainty</b>	
Uncertainty in $^{131}\text{I}$ transfer coefficient for commercial milk	---
Uncertainty in residence and diet history	--- <sup>d</sup>
Uncertainty in HTDS cohort mean dose <sup>e</sup>	---
Uncertainty in HTDS cohort dose variance <sup>e</sup>	---
Uncertainty in HTDS individual dose estimates	--- <sup>f</sup>
Stochastic variability of true dose	--- <sup>g</sup>
Uncertainty in thyroid doses from Nevada Test Site $^{131}\text{I}$	--- <sup>h</sup>

<sup>a</sup> +++ Potential for large overestimation, ++ moderate overestimation, + minor overestimation.

<sup>b</sup> --- Potential for large underestimation, -- moderate underestimation, - minor underestimation.

<sup>c</sup> HTDS reports 1,616 subjects with Nevada Test Site (NTS)  $^{131}\text{I}$  doses of less than 5 mGy; such small doses are not plausible for individuals consuming fresh sources of milk. Application of the NCI  $^{131}\text{I}$  Thyroid Dose/Risk Calculator (NCI 2003) for NTS Fallout (<http://ntsi131.nci.nih.gov/default.asp>) for residence locations and age groups for subjects in HTDS indicate doses that range from less than 3 to more than 20 times higher than the 5 mGy reference used by HTDS.

<sup>d</sup> Uncertainties in residence history and changes in dietary sources were not addressed in HTDS dose reconstruction.

<sup>e</sup> Uncertainties in the mean dose and dose variance were not reported for the HTDS cohort.

<sup>f</sup> Uncertainties on individual dose estimates too small given specified uncertainties for model covariates.

<sup>g</sup> Inter-individual stochastic variability of true dose only accounted for in HTDS by uncertainty assigned to dietary intake and thyroid dose conversion factor.

<sup>h</sup> Uncertainties in NTS  $^{131}\text{I}$  dose estimates were not addressed by HTDS.

the doses estimated for each single subject was 2.16. The range of individual dose GSDs was from 1.56 to 5.42.

We found that the dose uncertainties for individual subjects produced by the suite of Hanford dose reconstruction models are too low when compared with the uncertainty in the coefficients responsible for individual doses. The individual dose uncertainties reported for specific subjects in the HTDS are also low compared with the dose uncertainties produced by other modeling methodologies (NCI 1997; Stevens et al. 1992; Simon et al. 2006; Apostoaei 2005).

The uncertainty in the dose conversion factors for ingestion of  $^{131}\text{I}$  used to determine HTDS doses was described by a GSD of 2.0 (Snyder et al. 1994). Given that the dose conversion factor applied to a specific individual is highly correlated, even for an individual exposed in one or more years, the overall uncertainty in doses to specific members of the HTDS cohort should be characterized by a GSD of at least 2.0 or greater. Accounting for uncertainty in the dose conversion factors and for the uncertainty in other parts of the model (including the forementioned uncertainty in the  $F_m$  for commercial cow's milk), we estimate that the GSD in the doses for any member of the HTDS cohort should be described by a GSD of at least 2.3 or higher. However, more than 50% of the GSDs estimated in the HTDS for subject-specific doses are smaller than 2.3 (Davis et al. 2002). We believe that the reason for this apparent

suppression of uncertainty is the assumption of independence among uncertain model parameters when doses were obtained by summation over a series of days, months, and locations—conditions for which the uncertainty in model parameters is not independent.

For the dose distribution of an individual subject, a GSD of 2.3 or higher indicates high potential for dose misclassification, given that the uncertainty in each individual's dose will only be partially correlated with the uncertainty in the dose for other individuals in the HTDS cohort. Compared with the range of intra-individual dose uncertainty, the inter-individual variability of median doses in the HTDS was relatively low. The median doses for subjects ranged from 0.0029 to 2823 mGy, with a GSD of 2.69. Thus, the range of *uncertainty* about an individual's dose could span much of the entire range of the *variability* for the median individual doses for all study subjects.

### SIMULATIONS OF EXCESS RISK AND STATISTICAL POWER WITH DOSE UNCERTAINTIES

The HTDS authors noted that uncertainties in the individual dose estimates could reduce study power and conducted simulation analyses that showed the impact of these uncertainties on study power to be negligible (Davis et al. 2002). These analyses were expanded by

Stram and Kopecky (2003), with similar results. Because they treated most, if not all, uncertainty as Berkson, they could have substantially overestimated the power of the study. Though Stram and Kopecky acknowledged the importance of classical error, they did not include it in their analysis (2003, pp 416–417):

“An important qualifying issue, however, that merits some discussion here are errors in individual-level input data to the system, such as location, milk drinking habits, etc. In constructing a dosimetry system physicists tend to assume that these individual-level data are known without error, which may be very far from true in situations like the Hanford Thyroid Disease Study, which relied in part upon interview responses regarding personal behavior occurring more than 4 decades in the past. One approach to dealing with this issue from the point of view of the user of the system is to adopt a model for the errors in the input data and to perturb those data according to that model in repeated runs of the dosimetry system . . . Note that assuming a Berkson model for errors in the input data (in which the true input data is assumed to be distributed around the reported data independently for each individual) generally is not appropriate . . . A much better approach to dealing with errors in the input data is to adopt a classical error model in which the reported input data are distributed with mean equal to the truth. This leads to a classical “pull toward the center” of the distribution when the distribution of true input data around reported input data is constructed. . . . Perturbing . . . in this fashion reduces the variance . . . and hence reduces the estimate of study power appropriately” [emphasis added].

To explore the impact of Berkson and classical uncertainties on statistical power, we used the model of Reeves et al. (1998) and Mallick et al. (2002) to simulate true and calculated doses, with simulation assumptions described in the Appendix. The model has four variables: disease status  $Y$ , true dose  $X$ , calculated dose  $W$ , and a latent intermediate variable  $L$  between  $X$  and  $W$ :

$$Pr(Y = 1|X) = H[\alpha_{\text{sex}} + \log(1 + \beta X)] \quad (3)$$

$$\log(X) = \log(L) + U_b;$$

$$\log(W) = \log(L) + U_c;$$

$$H\{z\} = \text{logistic distribution function} = e^z/(1 + e^z);$$

$$U_b = \text{Berkson uncertainty};$$

$$U_c = \text{classical uncertainty}.$$

Simulations were performed with Berkson errors ranging between 70% and 100%, with correlation in uncertainties of 0% and 50%. The parameter values and supporting assumptions were selected to produce conditions similar to those for the dose-response analyses for thyroid neoplasia in the HTDS. Statistics were computed using likelihood methods. We restricted the ERR/Gy estimates and the upper end of confidence intervals to between 0.0 and 40.0. For hypothesis testing and confidence intervals, we used likelihood-ratio statistics; all tests were two-sided.

For these simulations, the true ERR/Gy was arbitrarily specified as 4.0, resulting in a statistical power of 0.80 when all uncertainty was unshared Berkson error. The simulation scenarios depended on the percentage of Berkson uncertainty and the common correlation in the shared Berkson uncertainties. For each of the eight scenarios, 500 simulated data sets were constructed. The mean number of disease cases for each scenario was computed by simulating samples of 3,000 subjects a total of 5,000 times and then averaging.

The results of the simulation experiment (Table 3) show that the simulated estimates of the number of cases and ERR per Gy are comparable to those reported in the HTDS, indicating that inferences regarding the effects of dose uncertainty on statistical power are relevant to the conditions of the HTDS dose-response analysis. The simulations indicate that shared (correlated) Berkson uncertainties lower statistical power over the spectrum of

**Table 3.** Results of simulation experiment<sup>a</sup> for effects of dose uncertainty on statistical power.

Correlation of Berkson errors	Percentage of uncertainty being Berkson	Mean number of cases	Mean ERR <sup>b</sup> Gy <sup>-1</sup>	Median ERR Gy <sup>-1</sup>	Mean Upper 95 <sup>th</sup> for ERR Gy <sup>-1</sup>	Median Upper 95 <sup>th</sup> for ERR Gy <sup>-1</sup>	Power
0.00	100%	40.7	5.7	3.9	23.1	40.0	0.75
	90%	40.6	4.8	3.2	20.6	40.0	0.64
	80%	40.7	3.4	2.4	17.0	24.2	0.55
	70%	40.7	3.0	2.0	14.8	20.8	0.43
0.50	100%	40.7	5.9	3.8	22.5	40.0	0.66
	90%	40.6	4.5	3.0	19.2	33.8	0.57
	80%	40.7	3.4	2.4	16.3	23.8	0.49
	70%	40.7	2.7	2.0	13.7	18.6	0.42

<sup>a</sup> See text and Appendix 1 for description of methods.

<sup>b</sup> ERR, excess relative risk.

combinations of Berkson and classical uncertainties. The greatest impact of shared uncertainty on power occurs when all uncertainty is Berkson, resulting in a reduction in power from 75% to 66% (Table 3). This is consistent with the results of Stram and Kopecky (2003) in their simulation of Berkson uncertainties.

The presence of classical uncertainties also lowers statistical power. For example, if all uncertainties are Berkson and unshared, the power is 75%; if the Berkson uncertainties are all unshared but 20% of the total uncertainty is classical, then the power drops to 55%; if Berkson errors are 50% unshared, then the power drops to 49%.

The upper end of the 95% confidence interval for ERR/Gy is not dramatically affected by the presence of shared Berkson uncertainties (Table 3). The presence of classical error leads to smaller upper confidence limits for the correctly specified measurement error-corrected estimator.

We also conducted a second set of simulations with conditions similar to those for the first set, except that the cumulative baseline risks per person were lower (for men, 0.0024, and for women, 0.0070) and the ERR was higher (6.5 per Gy). The results, not given here, paralleled the first simulation experiment, producing drops in power with increases in the proportion of shared Berkson uncertainty and with increases in the proportions of classical uncertainty.

Because it is possible that more than 20% of the uncertainty in HTDS doses estimates was classical, our simulations suggest that the HTDS dose-response analyses could have been underpowered. In any event, the intervals for the dose response presented by the HTDS do not incorporate uncertainty about the structure of the measurement error and hence are too narrow, particularly the upper ends of the confidence intervals for the estimates of the ERR per Gy. The extent of this problem can be studied with a more detailed analysis of the classical uncertainties in the HTDS dose estimates.

### ESTIMATES OF EXCESS RISK WITH IMPROVED ACCOUNTING FOR DOSE UNCERTAINTIES

Regardless of whether or not the HTDS was underpowered, it is possible to make inferences about the exposure-risk relation based on existing data. Specifically, accounting more thoroughly for the effects of dose uncertainty on estimating the thyroid cancer risk should raise the central values and widen the confidence intervals of the risk estimates reported by Davis et al. (2002, 2004). Even when dose uncertainty is ignored, the upper 95% confidence limit of the cumulative individual risk of

thyroid cancer per Gy reported by the HTDS is 0.017. This is equivalent to an ERR at 1 Gy of 4.2 [commensurate with a cumulative background risk of 0.006 per person for females and 0.002 for males (Davis et al. 2002)]. When uncertainty in the slope per Gy and cumulative background are taken into account, the ERR/Gy ranges from 0.0 to 4.6, with a central value of 0.48 (Table 1). An ERR per Gy of 4.6 is a value well within the confidence intervals of the ERR reported in other epidemiologic studies (Table 1) and produced by the preferred models of BEIR VII (NRC/NAS 2006).

Mallick et al. (2002) have shown that accounting for mixtures of Berkson and classical measurement errors will increase the upper confidence limits of ERR/Gy estimates by a factor of as much as two over those produced using mean estimates of individual doses and ignoring the presence of dose uncertainty. Increasing the upper confidence limit of the HTDS ERR at 1 Gy by factors of 2 to 4 (from 4.6 to 9.2, up to 18.4) to account for complex combinations of mixtures of systematic and random sources of dose uncertainty will cause the range of cancer risk estimates in the HTDS to overlap the central values of most ERR per Gy estimates published for children exposed to external radiation or fallout <sup>131</sup>I (Table 1).

The estimates of ERR per Gy in Table 1 have been adjusted using a probability distribution for the DDREF, in order to account for differences in the ERR per Gy between acute exposures to high energy gamma radiation (or fractionated exposures to x rays) and the ERR per Gy resulting from chronic, low dose-rate exposures to low-LET radiation, such as for exposure to <sup>131</sup>I. The probability distribution of the DDREF is taken from Land et al. (2003), and is specific for the radiogenic risk of thyroid cancer. Although not shown, similar results would have been obtained using the probability distribution of DDREF for all solid tumors, as was recently recommended by the BEIR VII Committee of the National Academy of Sciences (NRC/NAS 2006).

### STATISTICAL POWER FOR ALTERNATE ANALYSES

The HTDS conducted a variety of alternate analyses to determine whether their results could have been influenced by a variety of factors associated with dose estimation and statistical analyses. These included: 1) alternate definitions of disease outcome; 2) alternate dose-response models; 3) alternate sources of exposure data for individual subjects; 4) explorations of the impact of including or excluding participants who might have strong influences on dose-response relations; and 5) assessments of confounding and effect modification by

selected covariates for outcomes other than thyroid neoplasia (Davis et al. 2002, 2004). The HTDS authors determined that the conclusions from their primary analyses were not altered by the results of the alternate analyses (Davis et al. 2004).

The HTDS also conducted analyses that did not rely on estimated doses for individual subjects by using county of residence as a surrogate for dose (Davis et al. 2002). The results from these analyses did not change the results from the primary analyses, and the study authors concluded that this evidence provided some reassurance that exposure misclassification was not responsible for the results in their primary analyses (Davis et al. 2004).

Each of the alternate analyses was performed with the primary analytic dataset, or with a subset composed of fewer subjects than the original—resulting in the same problems with low statistical power that were identified for the primary analyses. Since ecologic exposure estimates are far more uncertain and variable than the individual doses used in the HTDS primary analyses, the statistical power for the alternate analyses would be expected to be even lower than for the primary analyses.

Consistency between negative findings from epidemiologic studies with low statistical power is not informative with regard to the issue of disease causation, as both the absence of a causal relation and the presence of type-two errors can explain the consistency equally well. Therefore, the negative results from the alternate HTDS analyses do not definitively support the primary findings. The same can be said for the consistency between the HTDS findings and those from a recent analysis of thyroid cancer mortality in Hanford-area counties (Boice et al. 2006)—an ecologic study with a high potential for exposure uncertainty, misclassification and low statistical power.

It is not surprising, therefore, that the alternate analyses produced statistically inconclusive results.

### PROPOSED EXPLANATIONS FOR HTDS NEGATIVE FINDINGS

Both the HTDS researchers and the BEIR VII Committee proposed possible explanations for why the HTDS results are different from those of the studies of persons exposed to  $^{131}\text{I}$  from the NTS and Chernobyl (Davis et al. 2002, 2004; NRC/NAS 2006). These include: (a) doses from short-lived radionuclides at Chernobyl and NTS, rather than  $^{131}\text{I}$ , were responsible for the elevated risks for thyroid cancer in studies of NTS and Chernobyl populations; (b) the prevalence of an iodine-deficient diet in Chernobyl populations was responsible for the elevated thyroid cancer risks from Chernobyl; (c) investigator bias explains the positive

findings in studies of children exposed to NTS fallout; and (d) children at Hanford were exposed to  $^{131}\text{I}$  at dose rates that were substantially lower than those for other exposed populations, and the cancer risk at these lower dose rates is less than that from higher dose rates.

The first three of these explanations were proposed with no supporting data and are contradicted by available evidence. Short-lived radionuclides were not important contributors to thyroid doses from NTS fallout; rather,  $^{131}\text{I}$  has been found to be the major source, and thyroid doses in the reports by Simon et al. (2006) and Lyon et al. (2006) were based on doses that were predominantly from  $^{131}\text{I}$  and  $^{133}\text{I}$ . For Chernobyl, the doses to children were predominantly from  $^{131}\text{I}$ , with external radiation and shorter-lived isotopes of iodine and other radionuclides contributing merely a few percent to the total dose (Gavrilin et al. 2004; Cardis et al. 2005; Minenko et al. 2006). There is no evidence to support the theory that the excess risk for thyroid cancer from  $^{131}\text{I}$  in Chernobyl populations is due exclusively to iodine-deficient diets, and recent findings indicate that the effects of iodine deficiency are independent of those from  $^{131}\text{I}$  (Cardis et al. 2005; Tronko et al. 2006).

Regarding investigator bias in the reports of Stevens et al. (1992) and Kerber et al. (1993), the reanalysis of these data included a blind review of all diagnoses by medical experts who concurred with most of the original diagnoses; most corrections in the reanalysis were to dose estimates for individual subjects (Lyon et al. 2006; Simon et al. 2006). Moreover, the finding of a dose-response relation based on estimated doses (which were unknown at the time of examination) is unlikely to have been influenced by medical examiners knowing where a subject lived and whether or not they had been diagnosed with thyroid cancer.

### WAS THE PATTERN OF HANFORD $^{131}\text{I}$ EXPOSURES UNIQUE?

The fourth proposed explanation for the differences between the results of the HTDS and those of NTS and Chernobyl cohorts is based on the untested hypothesis that there are differences between the biological effectiveness of various low-dose-rate exposures to ionizing radiation. The HTDS authors proposed that the  $^{131}\text{I}$  releases from Hanford occurred over longer time periods than those from the NTS or from Chernobyl, and that the thyroid cancer risk per unit dose from  $^{131}\text{I}$  at these lower dose rates was much less than the risk from exposures at higher dose rates. This explanation is based on two suppositions: the doses from Hanford  $^{131}\text{I}$  were actually delivered at dose rates that were substantially different

from those for the NTS and Chernobyl, and the difference in dose rates could produce the differences between excess risk estimates that were observed.

At Hanford, the  $^{131}\text{I}$  releases were nearly episodic in frequency, and the majority of the total dose to the thyroid glands of exposed persons was delivered over several months during 1945 (Heeb and Bates 1994). Because most of the subjects in the HTDS cohort received the greatest contribution to total individual dose from  $^{131}\text{I}$  releases that occurred in 1945, thyroid doses were actually delivered over time periods that were not much longer than those for Chernobyl or the NTS. By comparison, the majority of the thyroid dose for Chernobyl subjects was delivered over a period of about 6 weeks following the initial 10 days of releases that occurred beginning on 26 April 1986 (UNSCEAR 2000). For the NTS cohorts, doses were accumulated during a period of about four weeks subsequent to each deposition event, and many subjects were exposed to fallout from multiple atmospheric tests (Simon et al. 2006). Although there are differences between the low dose rates encountered at Hanford, the NTS and Chernobyl, these differences are not substantial.

We are aware of no evidence that directly supports the hypothesis that modest differences in low dose rates substantially influence thyroid cancer risks from  $^{131}\text{I}$  exposures in the ranges of thyroid doses reported for NTS, Chernobyl and Hanford. At no point in its report did the BEIR VII Committee (NRC/NAS 2006) conclude that prolonged exposures at very low dose rates result in lower risks per unit dose than chronic low dose-rate exposures at somewhat higher dose rates. Rather, the BEIR VII Committee applied the same distribution for the DDREF over the entire spectrum of low dose rates, including doses accumulated chronically and continuously over a lifetime.

Moreover, the BEIR VII committee did not conclude that the risk per unit dose for thyroid cancer from exposures to pure  $^{131}\text{I}$  would be lower than the risk from internal exposure to mixtures of  $^{131}\text{I}$  and other shorter-lived radionuclides or from fractionated or prolonged exposure to external sources of radiation. On the other hand, BEIR VII did acknowledge evidence for a higher radiobiological effectiveness on the order of a factor of two to three associated with exposure to low-energy photons or x rays. These differences are small compared to the order-of-magnitude difference between point estimates from the HTDS dose-response models for thyroid cancer and those reported for other studies (Table 1).

The HTDS low-dose-rate hypothesis has little support from the consensus of experts or from epidemiologic evidence. Although data from the HTDS might be used in re-analyses to explore whether the risks for subjects

who received doses at very low dose rates are different from those for subjects who received doses delivered over shorter time intervals, it is unlikely that such analyses would have adequate statistical power.

## CONCLUSION

We have provided evidence that the negative results of the HTDS could be due to inadequate statistical power due to the presence of large and complex uncertainties in the modeled doses for individual subjects. Although HTDS researchers have addressed this possibility to some extent (Stram and Kopecky 2003; Kopecky et al. 2004), there is a need for additional analyses. Further investigations should address the effects on statistical power or interval estimates of the ERR/Gy from shared and unshared Berkson errors mixed with classical errors, combined with varying potential degrees of systematic bias within the dose estimates for each individual member of the HTDS cohort.

Even if there is no consensus that low statistical power best explains the negative results, the incomplete accounting for complex combinations of uncertainties about the error structure of individual dose estimates has resulted in interval estimates on the dose-response relationships for specific thyroid disease endpoints that are too narrow. These should be adjusted appropriately.

Similar conclusions were reached by Kopecky et al. (2006), who addressed the effect of classical errors in dose estimates on the ERR/Gy for thyroid cancer in childhood cohorts in Russia who were exposed to Chernobyl  $^{131}\text{I}$ . By incorporating the effect of classical errors, the point estimate of the ERR/Gy nearly tripled over the case where uncertainty in dose estimation was ignored. In addition, the 95% interval about the point estimate of the ERR/Gy increased markedly.

Until the issues we have identified are resolved, it is premature to conclude that excess thyroid cancer risks in the Hanford cohort are lower than those that have been estimated from other epidemiologic studies. We recommend that in the interim, estimates of excess risk for persons exposed to  $^{131}\text{I}$  from Hanford combine data from other epidemiologic studies with those from the HTDS. This task will require a more detailed analysis of the effect of errors in dosimetry on the interval estimates of the ERR per Gy, with methods of Monte Carlo or Bayesian bias analyses, as suggested by Greenland (2005).

Although we have addressed only thyroid neoplasia, our concerns are also relevant to the other disease endpoints evaluated in the HTDS, as are our recommendations for estimating excess risks of these thyroid

diseases. We hope that our paper will foster continuing debate and discussion of these issues.

*Acknowledgments*—We would like to thank Duncan Thomas of the University of Southern California for many helpful discussions and comments.

**Disclosures:** The authors have provided scientific litigation support on the behalf of plaintiffs in lawsuits involving radiation exposure to the public from Hanford. This paper was developed to involve the wider scientific community in discussions of the HTDS results. The manuscript is solely the work of the authors; it has not been reviewed or edited by others involved in Hanford litigation; nor have the authors received funds for its preparation.

## REFERENCES

- Apostoaie AI. Testing prediction capabilities of a <sup>131</sup>I terrestrial transport model by using measurements collected at the Hanford nuclear facility. *Health Phys* 88:439–458; 2005.
- Boice JD, Mumma MT, Blot WJ. Cancer mortality among populations residing in counties near the Hanford site, 1950–2000. *Health Phys* 90:431–445; 2006.
- Cardis E, Kesminiene A, Ivanov V, Malakhova I, Shibata Y, Khrouch V, Drozdovitch V, Maceika E, Zvonova I, Vlassov O, Bouville A, Goulko G, Hoshi M, Abrosimov A, Anoshko J, Astakhova L, Chekin S, Demidchik E, Galanti R, Ito M, Korobova E, Lushnikov E, Maksoutov M, Masyakin V, Nerovnia A, Parshin V, Parshkov E, Piptsevich N, Pichera A, Polyakov S, Shabeka N, Suonio E, Tenet V, Tsyb A, Yamashita S, Williams D. Risk of thyroid cancer after exposure to <sup>131</sup>I in childhood. *J Nat Cancer Inst* 97:724–732; 2005.
- Carroll RJ, Ruppert D, Stefanski LA, Crainceanu CM. Measurement error in nonlinear models: a modern perspective. Boca Raton, FL: Chapman and Hall CRC Press; 2006.
- Davis S, Kopecky KJ, Hamilton TE, Onstad LE, King BL, Saporito MS. Hanford thyroid disease study: Final report. 2002. Available at: <http://www.cdc.gov/nceh/radiation/hanford/htdsweb/>. Accessed 27 Oct 2005.
- Davis S, Kopecky KJ, Hamilton TE, Onstad LE, King BL, Saporito MS. Thyroid neoplasia, autoimmune thyroiditis, and hypothyroidism in persons exposed to iodine-131 from the Hanford Nuclear Site. *JAMA* 292:2600–2613; 2004.
- Farris WT, Napier BA, Ikenberry TA, Simpson JC, Shipler DB. Atmospheric pathway dosimetry report, 1944–1992. Richland, WA: Battelle Pacific Northwest Laboratories; PNWD-2228 HEDR; 1994.
- Gavrillin Y, Khrouch V, Shinkarev S, Drozdovitch S, Minenko V, Shemiakina E, Ulanovsky A, Bouville A, Anspaugh L, Voillequé P, Luckyanov N. Individual thyroid dose estimation for a case-control study of Chernobyl-related thyroid cancer among children of Belarus—Part I: <sup>131</sup>I, short-lived radioiodines (<sup>132</sup>I, <sup>133</sup>I, <sup>135</sup>I), and short-lived radiotelluriums (<sup>131m</sup>Te and <sup>132</sup>Te). *Health Phys* 86:565–585; 2004.
- Gilbert ES, Tarone R, Bouville A, Ron E. Thyroid cancer rates and <sup>131</sup>I doses from Nevada atmospheric nuclear bomb tests. *J Nat Cancer Inst* 90:1654–1660; 1998.
- Greenland S. Multiple-bias modeling for analysis of observational data (with discussion). *J Roy Statist Soc, Series A* 168:267–308; 2005.
- Heeb CM, Bates DJ. Radionuclide releases to the Columbia River from Hanford operations, 1944–1971. Richland, WA: Battelle Pacific Northwest Laboratories; PNWD2223-HEDR; 1994.
- Hoffman FO. Peer review of HEDR uncertainty and sensitivity analyses plan. Hanford environmental dose reconstruction project. Richland, WA: Battelle Pacific Northwest Laboratories; PNWD-2162 HEDR; 1993.
- Hoffman FO. How large can uncertainty be when models take the place of measurements? In: Ron E, Hoffman FO, eds. Uncertainties in radiation dosimetry and their impact on dose-response analysis. Bethesda, MD: NCI; NIH 99-4541; 1999.
- Institute of Medicine/National Research Council. Exposure of the American people to iodine-131 from Nevada nuclear bomb tests. Review of the National Cancer Institute report and public health implications. Washington, DC: National Academy Press; 1999.
- Ivanov VK, Gorski AI, Pitkevitch VA, Tsyb AF, Cardis E, Storm H. Risk of radiogenic thyroid cancer in Russia following the Chernobyl accident. In: Thomas G, Karaoglou A, Williams ED, eds. Radiation and thyroid cancer. River Edge, NJ: World Scientific Publishing; 1999.
- Jacob P, Kenigsberg Y, Zvonova I, Goulko G, Buglova E, Heidenreich WF, Golovneva A, Bratilova AA, Dorzdovitch V, Kruk J, Pochtennaja GT, Balonov M, Demidchik EP, Paretzke HG. Childhood exposure due to the Chernobyl accident and thyroid cancer risk in contaminated areas of Belarus and Russia. *Br J Cancer* 80:1461–1469; 1999.
- Jacob P, Kenigsberg J, Likhtarev I, Tronko M, Shinkarev S, Meckbach R, Demidchik Y, Kovgan L, Bogdanova T, Ulanovski A, Gavrillin Y, Buglova E, Vavilov S, Schotola C, Kruk J, Chepurniy M, Prohl G. Thyroid exposure of Belarusian and Ukrainian children due to the Chernobyl accident and resulting thyroid cancer risk. Final report of BfS project StSch4240. Neuherberg: GSF—Forschungszentrum für Umwelt und Gesundheit, GmbH; 2004.
- Kerber RA, Till JE, Simon SL, Lyon LL, Thomas DC, Preston-Martin S, Ralison ML, Lloyd RD, Stevens W. A cohort study of thyroid disease in relation to fallout from nuclear weapons testing. *JAMA* 270:2076–2082; 1993.
- Kopecky KJ, Davis S, Hamilton TE, Saporito MS, Onstad LE. Estimation of thyroid radiation doses for the Hanford thyroid disease study: results and implications for statistical power of the epidemiologic analyses. *Health Phys* 87:15–32; 2004.
- Kopecky KJ, Onstad LE, Hamilton TE, Davis S. Thyroid ultrasound abnormalities in persons exposed during childhood to <sup>131</sup>I from the Hanford nuclear site. *Thyroid* 15:604–613; 2005.
- Kopecky KJ, Stepanenko NR, Rivkind N, Voilleque P, Onstad L, Shakhtarin V, Parshkov E, Kulikov S, Lushnikov E, Abrosimov A, Troshin V, Romanova G, Doroschenko V, Proshin A, Tsyb A, Davis S. Childhood thyroid cancer, radiation dose from Chernobyl, and dose uncertainties in Bryansk Oblast, Russia: a population-based case-control study. *Rad Res* 166:367–374; 2006.
- Land C, Gilbert E, Smith J, Hoffman FO, Apostoaie IA, Thomas BA, Kocher DC. Report of the NCI-CDC working group to revise the 1985 NIH radioepidemiological tables. Bethesda, MD: NIH/NCI; 2003.
- Lindberg S, Karlsson P, Arvidsson B, Holmberg E, Lundberg L, Wallfren A. Cancer incidence after radiotherapy for skin hemangioma during infancy. *Acta Oncologica* 34:735–740; 1995.
- Lubin JH, Schafer DW, Ron E, Stovall M, Carroll RJ. A reanalysis of thyroid neoplasms in the Israeli tinea capitis

- study accounting for dose uncertainties. *Radiat Res* 161:359–368; 2004.
- Lundell M, Hakulinen T, Holm L-E. Thyroid cancer after radiotherapy for skin hemangioma in infancy. *Radiat Res* 140:334–339; 1994.
- Lyon JL, Alder SC, Bishop Stone M, Scholl A, Reading JC, Holubkov R, Sheng X, White GL Jr, Hegmann KT, Anspaugh L, Hoffman FO, Simon SL, Thomas B, Carroll R, Meikle AW. Thyroid disease associated with exposure to the Nevada Nuclear Weapons Test Site radiation: A reevaluation based on corrected dosimetry and examination data. 2006. *Epidemiology* 17:604–614; 2006.
- Mallick B, Hoffman FO, Carroll RJ. Semiparametric regression modeling with mixtures of Berkson and classical error, with application to fallout from the Nevada Test Site. *Biometrics* 58:13–20; 2002.
- Minenko VF, Ulanovsk AV, Drozdovitch VV, Shemiakina EV, Gavrilin YI, Khrouch VT, Shinkarev SM, Voillequé PG, Bouville A, Anspaugh LR, Luckyanov NL. Individual thyroid dose estimates for a case-control study of Chernobyl-related thyroid cancer among children of Belarus—Part II. Contributions from long-lived radionuclides and external radiation. *Health Phys* 90:312–327; 2006.
- National Cancer Institute. Estimated exposures and thyroid doses received by the American people from iodine-131 in fallout following Nevada atmospheric nuclear bomb tests: a report from the National Cancer Institute. Bethesda, MD: U.S. Department of Health and Human Services; 1997.
- National Cancer Institute. <sup>131</sup>I thyroid dose/risk calculator for Nevada Test Site fallout. 2003. Available at <http://ntsi131.nci.nih.gov/>. Accessed 28 Oct 2005.
- National Research Council/National Academies of Science. Review of the Hanford thyroid disease study draft final report. Washington, DC: National Academy Press; 2000.
- National Research Council/National Academies of Science. Health risks from exposure to low levels of ionizing radiation—BEIR VII (Phase 2). A report of the Board on Radiation Effects Research. Washington, DC: National Academies Press; 2006.
- Reeves GK, Cox DR, Darby SC, Whitley E. Some aspects of measurement error in explanatory variables for continuous and binary regression models. *Statistics in Medicine* 17:2157–2177; 1998.
- Ron E, Lubin JH, Shore RE, Mabuchi K, Modan B, Pottern LM, Schneider AB, Tucker MA, Boice JD Jr. Thyroid cancer after exposure to external radiation: a pooled analysis of seven studies. *Radiat Res* 141:259–277; 1995.
- Rothman KJ, Greenland S. *Modern epidemiology*, 2nd edition. Philadelphia: Lippincott Williams & Wilkins; 1998.
- Schafer DW, Gilbert ES. Some statistical implications of dose uncertainty in radiation dose-response analyses. *Rad Res* 166:303–312; 2006.
- Shakhtarin VV, Tsyb A, Stepanenko VF, Orlov MY, Kopecky KJ, Davis S. Iodine deficiency, radiation dose, and the risk of thyroid cancer among children and adolescents in the Bryansk region of Russia following the Chernobyl power station accident. *Int J Epidemiol* 32:584–591; 2003.
- Shore R. Issues and epidemiological evidence regarding radiation-induced thyroid cancer. *Radiat Res* 131:98–111; 1992.
- Shore R, Xue X. Comparative thyroid cancer risk of childhood and adult radiation exposure and estimation of lifetime risk. In: Thomas G, Karaoglou A, Williams ED, eds. *Radiation and thyroid cancer*. Brussels-Luxembourg: World Scientific Publishing Ltd; 1999: 491–498.
- Simon SI, Anspaugh LR, Hoffman FO, Scholl AE, Stone MB, Thomas BA, Lyon JL. 2004 update of dosimetry for the Utah Thyroid Cohort Study. *Radiat Res* 165:208–222; 2006.
- Snyder FS, Farris WT, Napier BA, Ikenberry TA, Gilbert RO. Parameters used in the environmental pathways and radiological dose modules (DESCARTES, CIDER, and CRD codes) of the Hanford Environmental Dose Reconstruction Integrated Codes (HEDRIC). Richland, WA: Battelle Pacific Northwest Laboratories; PNWD-2023 HEDR Rev. 0; 1994.
- Stevens W, Till JE, Thomas DC, Lyon JL, Kerber RA, Preston-Martin S, Simon SL, Rallison ML, Lloyd RD. Assessment of leukemia and thyroid disease in relation to fallout in Utah. Report of a cohort study of thyroid disease and radioactive fallout from the Nevada Test Site. Salt Lake City, UT: University of Utah Press; 1992.
- Stram DO, Kopecky KJ. Power analysis of epidemiologic studies of radiation-related disease risk when dose estimates are based on a complex dosimetry system with an application to the Hanford thyroid disease study. *Radiat Res* 160:408–417; 2003.
- Thompson DE, Mabuchi K, Ron E, Soda M, Tokunaga M, Ochikubo S, Sugimoto S, Ikeda T, Terasaky M, Izumi S, Preston DL. Cancer incidence in atomic bomb survivors. Part II: Solid tumors, 1958–1987. *Rad Res* 137:S17–S67; 1994.
- Tronko MD, Howe GR, Bogdanova IT, Bouville AC, Epstein OV, Brill AB, Likhtarev IA, Fink DJ, Markov VV, Greenebaum E, Olijnyk VA, Masnyk IJ, Shpak VM, McConnell RJ, Tereshchenko VP, Robbins J, Zvinchuk OV, Zablotska LB, Hatch M, Luckyanov NK, Ron E, Thomas TL, Voilleque PG, Beebe GW. A cohort study of thyroid cancer and other thyroid diseases after the Chernobyl accident: thyroid cancer in Ukraine detected during first screening. *J Nat Cancer Inst* 98:897–903; 2006.
- United National Scientific Committee on the Effects of Atomic Radiation. Sources and effects of ionizing radiation. UNSCEAR 2000 report to the general assembly, with scientific annexes. Volume I: Sources. New York: United Nations; 2000.

## APPENDIX

### Assumptions for simulations of statistical power

In our simulations, the Berkson and classical uncertainties are normally distributed with mean zero and standard deviations  $\sigma_b$  and  $\sigma_c$ , respectively. The latent intermediate variable  $L$  is lognormally distributed. The parameter values chosen for the simulations are as follows:

a. The number of subjects for the simulations is 3,000, with equal numbers of males and females—similar to the composition of the population in the HTDS dose-response analyses (Davis et al. 2004).

b. For simulations, the mean calculated dose in the log scale is  $\log(0.10)$  Gy, an estimate based on the log of the median dose in the HTDS cohort (0.097 Gy) (Davis et al. 2004). The standard deviation of the HTDS calculated doses is 0.224 Gy (Davis et al. 2004); because

there are some major outliers in the dose distribution, we assumed a standard deviation of 0.175 Gy.

c. The standard deviation of calculated doses in the log scale can be calculated in two ways. The mean of the HTDS calculated doses is 0.174 Gy (Davis et al. 2002). Assuming log-normality, if  $s$  is the standard deviation of calculated dose in the log scale, then the mean of calculated doses divided by the median calculated dose is  $\exp(\sigma^2/2)$ , and  $\sigma = 1.09$ . Alternatively, the variance of the calculated doses divided by the square of the median of the calculated dose is  $\exp(2\sigma^2) - \exp(\sigma^2)$ , suggesting a value of  $\sigma = 0.93$ . As a compromise between these two values of  $\sigma$ , we selected an intermediate value:  $\sigma = \log(2.7) = 0.99$ .

d. For simulations, the standard deviation of uncertainties in the log scale is  $\log(2.3)$ . In the HTDS report (Davis et al. 2002), the ratio of the upper 95<sup>th</sup> percentile dose to the median dose for each individual subject is approximately 4.0. If  $\mu$  is the mean of the calculated dose in the log scale, and if  $\kappa$  is the standard deviation of the uncertainty in the log scale, then the median calculated dose is approximately  $\exp(\mu)$ , while the 95<sup>th</sup> percentile for the individual is approximately  $\exp(\mu + 1.645 \kappa)$ . This suggests that  $\exp(1.645 \kappa) = 4.0$ , or that  $\kappa = \log(2.32)$ .

e. The percentages of error assigned as Berkson are 100%, 90%, 80%, and 70%. This range is higher than for the analyses with Bayesian models developed by the HTDS (Davis et al. 2002). The correlation of shared Berkson errors is 0.0 and 0.5.

f. The baseline thyroid neoplasia risk for men is  $0.0049 - \Delta$ , and for women,  $0.0098 - \Delta$ . The baseline risks for men and women when all error was Berkson were 0.0049 and 0.0098, respectively. When classical error was allowed for, we adjusted the risks slightly by subtracting a quantity  $\Delta$ , chosen to keep the mean

number of cases in all simulations at approximately 44. For example, when 80% of the error was Berkson, the baseline risks for men and women were 0.0052 and 0.0103, respectively.

g. Because this is a simulation experiment in which true doses and calculated doses are generated in the log scale, a decision was necessary as to what the calculated doses were in the arithmetic scale of the Gray. We took the mean, assuming that all uncertainties were Berkson. Thus, for an individual, if his or her calculated dose in the log scale is  $Z_L$ , and the standard deviation of the uncertainty is  $\kappa$ , then the calculated dose in the Gray scale is  $\exp(Z_L + \kappa^2/2)$ .

h. In the model of Reeves et al. (1998) and Mallick et al. (2002), there is a latent variable in the log scale that the latter authors call a "latent intermediate"  $L$ . In principle, this is a normal random variable with mean  $m_L$  and standard deviation  $s_L$ . The difficulty with this is that if one exponentiates  $L$ , one can get extremely high true doses in the scale of the Gray. In order to avoid this happening,  $L$  was not allowed to be larger than  $m_L + 2 s_L$ .

i. Estimates, test statistics, and confidence intervals were computed using likelihood methods. For estimation, we restricted the ERR/Gy and its confidence interval to be between 0.0 and 40.0. The true ERR/Gy was taken to equal 4.0 so that the power when all uncertainty was unshared Berkson was approximately 0.80.

j. Scenarios depended on the percentage of uncertainty that is Berkson and the common correlation in the shared Berkson uncertainties. For each of the eight scenarios, 500 simulated data sets were constructed.

k. The mean number of disease cases for each scenario was computed by simulating samples of size  $n = 3,000$  a total of 5,000 times, and then averaging. ■ ■