

## Thyroid Cancer Following Scalp Irradiation: A Reanalysis Accounting for Uncertainty in Dosimetry

Daniel W. Schafer,<sup>1,\*</sup> Jay H. Lubin,<sup>2</sup> Elaine Ron,<sup>2</sup> Marilyn Stovall,<sup>3</sup> and Raymond J. Carroll<sup>4</sup>

<sup>1</sup>Department of Statistics, Oregon State University,  
Corvallis, Oregon 97331-4606, U.S.A.

<sup>2</sup>Biostatistics Branch and Radiation Epidemiology Branch,  
Division of Cancer Epidemiology and Genetics, National Cancer Institute,  
Executive Plaza South, 6120 Executive Boulevard, MSC 7368,  
Bethesda, Maryland 20892-7244, U.S.A.

<sup>3</sup>Department of Radiation Physics, University of Texas M. D. Anderson Cancer Center,  
1515 Holcombe Boulevard, Houston, Texas 77030, U.S.A.

<sup>4</sup>Department of Statistics, Texas A&M University, College Station, Texas 77843-3143, U.S.A.

\* *email:* schafer@stat.orst.edu

**SUMMARY.** In the 1940s and 1950s, over 20,000 children in Israel were treated for tinea capitis (scalp ringworm) by irradiation to induce epilation. Follow-up studies showed that the radiation exposure was associated with the development of malignant thyroid neoplasms. Despite this clear evidence of an effect, the magnitude of the dose–response relationship is much less clear because of probable errors in individual estimates of dose to the thyroid gland. Such errors have the potential to bias dose–response estimation, a potential that was not widely appreciated at the time of the original analyses. We revisit this issue, describing in detail how errors in dosimetry might occur, and we develop a new dose–response model that takes the uncertainties of the dosimetry into account. Our model for the uncertainty in dosimetry is a complex and new variant of the classical multiplicative Berkson error model, having components of classical multiplicative measurement error as well as missing data. Analysis of the tinea capitis data suggests that measurement error in the dosimetry has only a negligible effect on dose–response estimation and inference as well as on the modifying effect of age at exposure.

**KEY WORDS:** Berkson measurement error; Dose uncertainty; Dosimetry; Errors-in-variables; Likelihood; Measurement error; Missing data; Person-year tables; Poisson regression; Regression calibration; Survival analysis.

### 1. Introduction

Studies of the late effects of irradiation on a cohort of nearly 11,000 of the Israeli children irradiated for tinea capitis (scalp ringworm) showed an increasing risk of thyroid cancer associated with increasing dose of radiation to the thyroid gland (Modan, Ron, and Werner, 1977; Ron et al., 1989, 1995). The presence of an association has also been found in other studies on different populations (see Ron et al., 1995), but the surprising steepness of the dose–response relationship from the tinea capitis study has been called into question, particularly in light of the inaccuracies in the individual estimated doses. We reanalyzed the data to estimate relative risk models, this time accounting for uncertainties in thyroid radiation dose. The main epidemiological result is that the conclusions did not change appreciably. This article is concerned with the statistical approach and issues associated with adjustment for dose uncertainty in this problem.

The approach is necessarily customized to the particular

intricacies of this data problem. It is the illustration of the customization and the handling of departures from ideal situations that we believe are instructive. The most important component is the correct modeling of the multiple sources of dose uncertainty, especially the correct determination of uncertainties as classical measurement errors or Berkson errors. Of some statistical novelty is the use of a model that incorporates external prediction data, meaning data from a separate study that is used to formulate and estimate a model for predicting dose from predictor variables. This results in a Berkson uncertainty from using an expected dose in place of an individual's actual dose and, additionally, a classical uncertainty due to the sampling error in estimating unknown coefficients in the prediction model.

The gist of our statistical methodology is likelihood analysis for an induced hazard function given the available variables for predicting dose using both the primary data set and the secondary external prediction data set. Sensitivity

analyses were used to explore the consequences of inaccuracies in the parts of the input that were necessarily subjective. Related work on statistical tools for inference from censored survival times in the presence of covariate measurement error includes Pepe, Self, and Prentice (1989), Pierce, Stram, and Vaeth (1990), Pierce et al. (1992), Nakamura (1992), Hughes (1993), Thomas, Stram, and Dwyer (1993), Lubin, Boice, and Samet (1995), Hu, Tsiatis, and Davidian (1998), and Ron and Hoffman (1999).

The implementation is clouded by problems of missing data, different types of dose predictors for different subjects, different degrees of dose uncertainty for different subjects, and sources of dose uncertainty that require subjective assessment. Section 2 presents the basic statistical approach in general terms, stripped of the complicating features. Section 3 details the information available for dosimetry and the statistical models used to incorporate the uncertainty. Section 4 puts the pieces from Sections 2 and 3 together and illustrates the computational form of the analysis. The numerical results, including sensitivity analyses, are discussed in Section 5. Section 6 gives concluding remarks.

## 2. Overview of the Statistical Modeling and Analysis

### 2.1 A Model of Interest for the Hazard Function

The study population consists of 10,834 children who received x-ray therapy between 1948 and 1960, 10,834 nonirradiated population-matched controls, and 5392 nonirradiated tinea-free siblings. The rates of malignant thyroid tumors by 1986 were 4.0 per 1000 persons in the irradiated subjects and 1.0 per 1000 persons in the nonirradiated subjects.

Let  $T$  be a random variable representing the elapsed time after entry into the study until onset of thyroid cancer. Of interest in this article is inference about models of the following form for the age-specific thyroid cancer rate (hazard function of  $T$ ) as a function of total radiation dose to the thyroid,  $D$ , and additional covariates,  $X$ :

$$h(t | X, D, \mathcal{B}) = \exp \left\{ X_{\text{br}}^{\text{T}}(t) \beta_{\text{br}} \right\} \left[ 1 + D \exp \left\{ \beta_{\text{dr}} + X_{\text{em}}^{\text{T}}(t) \beta_{\text{em}} \right\} \right]. \quad (1)$$

In (1),  $X = X(t) = (X_{\text{br}}, X_{\text{em}})$ , where  $X_{\text{br}}$  is a vector of those covariates associated with background rate (hence br), such as time since exposure, sex, attained age, and place of origin. Let  $X_{\text{em}}$  be a vector of those covariates that modify the radiation dose-response, i.e., effect modifiers (hence em) such as sex and age at exposure. In what follows, we will suppress the notational dependence of  $X$  on  $t$ . The parameters  $\mathcal{B} = (\beta_{\text{br}}, \beta_{\text{em}}, \beta_{\text{dr}})$  are associated with  $(X_{\text{br}}, X_{\text{em}}, D)$ , and hence control background rate, effect modifiers, and dose response (hence dr). Model (1) is commonly used for carcinogenic effects of low doses of radiation, supported on both theoretical and empirical grounds (Ron et al., 1995).

To make inferences, Ron et al. (1989, 1995) treated the responses of the irradiated subjects as independent of those of the controls, i.e., they ignored the matching of irradiated subjects with their nonirradiated, matched controls and with their nonirradiated siblings. Ignoring the first dependence is justified by the use of the matching variables (age, sex, and place of origin) as covariates in the model. Ignoring the second

is not so justified, but since there were only six nonirradiated siblings who developed thyroid cancer, none of whom had a corresponding irradiated sibling who also developed thyroid cancer, it is unlikely that correctly accounting for sibling dependence, which would greatly complicate the analysis, will make any difference. We therefore will also proceed as if responses of all subjects, conditional on dose and other covariates, are mutually independent.

If the available doses are treated as exact, then standard techniques for relative risk regression from censored survival times, such as partial likelihood analysis of Cox regression models or likelihood analysis based on the Poisson subject-years method, could be used to make inferences about  $\mathcal{B}$  in (1) (cf., Breslow et al., 1983). If interest is in the regression on true dose, though, biases may result in estimates of  $\mathcal{B}$  from using the imprecise estimates of true dose (cf., Carroll, Rupert, and Stefanski, 1995).

Postponing the details of the dosimetry until the next section, we shall describe our statistical approach that accounts for dose uncertainty in general terms now. Let  $\mathcal{W}$  be a vector of variables that are available for predicting a child's total thyroid dose. For now, think of these as number of x-ray treatments received, type of x-ray machine used on each application, machine settings on each application, and age of the child on each application. The dose estimates used in the original analyses of the tinea capitis subjects were obtained, essentially, by applying a formula for predicting dose from  $\mathcal{W}$ . The formula was based on results from phantom studies (in which anthropomorphic phantoms, with dosimeters placed at the "thyroid," are exposed to x-rays). Our procedure considers the induced hazard function of  $T$  conditional on the observable variables  $\mathcal{W}$  and  $X$ .

### 2.2 Induced Hazard Function

Assume that  $T$  and  $\mathcal{W}$  are conditionally independent, given  $X$  and  $D$ , i.e.,

$$h(t | D, X, \mathcal{W}, \mathcal{B}) = h(t | D, X, \mathcal{B}), \quad (2)$$

which means that, given true dose and  $X$ , time to thyroid cancer is independent of the dose-predictor variables. Then, as shown in the Appendix,

$$h(t | X, \mathcal{W}, \mathcal{B}, \Theta) = E \{ h(t | D, X, \mathcal{B}) | T \geq t, X, \mathcal{W}, \Theta \}, \quad (3)$$

where  $\Theta$  represents the unknown parameters in the distribution of  $D$  given  $X$  and  $\mathcal{W}$ . Prentice (1982) showed this to be true for the special case that  $\mathcal{W}$  is a univariate measurement of  $D$ . In our setting,  $\mathcal{W}$  is a vector of dose-predictor variables.

Since the risk of disease is small, the approximation to (3) obtained by dropping the condition  $T \geq t$  should be adequate (as discussed in Pepe et al., 1989), in which case, the hazard based on the observable covariates  $X$  and  $\mathcal{W}$  is

$$h(t | X, \mathcal{W}, \mathcal{B}, \Theta) = E \{ h(t | D, X, \mathcal{B}) | X, \mathcal{W}, \Theta \} = h \{ t | E(D | X, \mathcal{W}, \Theta), X, \mathcal{B} \},$$

where the last equality holds because the assumed hazard function is linear in dose. Thus, the parameters of interest are in the hazard function given the observable variables,

$$h(t | X, \mathcal{W}, \mathcal{B}, \Theta) = \exp \left( X_{\text{br}}^{\text{T}}(t) \beta_{\text{br}} \right)$$

$$\times \left[ 1 + E(D | X, \mathcal{W}, \Theta) \exp \left\{ \beta_{\text{dr}} + X_{\text{em}}^{\text{T}}(t) \beta_{\text{em}} \right\} \right]. \quad (4)$$

If, e.g.,  $\Theta$  is known and  $E(D | X, \mathcal{W}, \Theta)$  can be calculated for each subject in the tinea study group, then parameters may be estimated with standard methods but with unknown doses replaced by their expectations given  $X$  and  $\mathcal{W}$ . In general, the technique of using usual methods but with the unknown doses replaced by their expectations given available variables is known as regression calibration (Carroll et al., 1995).

If  $\Theta$  were known, then the parameters in the relative risk portion of (4), i.e., the part in brackets, could be estimated either by the partial likelihood analysis of the Cox regression model or by the subject-years method, which is based on Poisson likelihood calculations of cancer occurrences tabulated over intervals of time after entry into the study. We shall focus on the latter because we wish to use an exact likelihood function. In that way, we can combine the likelihood associated with the tinea subjects with the likelihood associated with the phantom studies in order to simultaneously estimate  $\mathcal{B}$  (the parameters in the relative risk regression) and  $\Theta$  (the parameters in the model for predicting dose from the dose predictor variables). Inferences about  $\mathcal{B}$  will automatically account for the uncertainty in estimating  $\Theta$  without any need for *ad hoc* adjustments.

### 2.3 The Combined Likelihood Function

The density function for  $T_i$  for the  $i$ th member of the tinea study group is

$$f(t | D_i^\Theta, X_i, \mathcal{B}, \Theta) = h(t | D_i^\Theta, X_i, \mathcal{B}) \exp \left\{ - \int_0^t h(u | D_i^\Theta, X_i, \mathcal{B}) du \right\},$$

where  $D_i^\Theta$  is an abbreviation for  $E(D_i | X_i, \mathcal{W}_i, \Theta)$  and  $h(t | D_i^\Theta, X_i, \mathcal{B})$  is the hazard function in (4) (cf., Cox and Oakes, 1984, Section 2.2). Assuming independence, the log-likelihood function from the tinea subjects is given by

$$\ell_{\text{tinea}}(\mathcal{B}, \Theta) = \sum_i \left[ \nu_i \log \left\{ h(T_i | D_i^\Theta, X_i, \mathcal{B}) \right\} - \int_0^{S_i} h(u | D_i^\Theta, X_i, \mathcal{B}) du \right],$$

where  $\nu_i$  is an indicator variable for uncensored observations and  $S_i$  is the minimum of  $T_i$  and the censoring time (Cox and Oakes, 1984, Section 3.2).

The subject-years or Poisson approach offers relatively simple calculations by tabulation according to various states that the subjects pass through during the course of observation. For a given fixed value of  $\Theta$ , the tinea data may be cross-classified according to  $J$  states formed by all combinations of various categories of time since exposure and explanatory variables such as sex, country of origin, expected dose, and age at first exposure. With the inconsequential assumption that the covariables  $(X, D^\Theta)$  take on constant

values within states, the log likelihood reduces to

$$\ell_{\text{tinea}}(\mathcal{B}, \Theta) = \sum_{j=1}^J \left[ O_j \log \left\{ h(t | D_j^\Theta, X_j, \mathcal{B}) \right\} - h(t | D_j^\Theta, X_j, \mathcal{B}) E_j \right], \quad (5)$$

where  $O_j$  is the observed number of cancers and  $E_j$  is the person-years of observation in state  $j$  (see Breslow et al., 1983). The actual value used for  $D_j^\Theta$  for a given  $\Theta$  is the person-years weighted average of expected dose for all individuals  $i$  observed in state  $j$ .

One approach is to estimate  $\Theta$  from the phantom data, then treat it as known for maximum likelihood estimation of  $\mathcal{B}$  from (5). We call this regression calibration. In this, however, the covariate  $D_j^{\hat{\Theta}}$  is an imprecise estimate of the explanatory variable of interest,  $D_j^\Theta$ . Writing  $D_j^{\hat{\Theta}} = D_j^\Theta + \epsilon_j$ , where  $\epsilon_j$  represents the imprecision due to sampling variability in estimating  $\Theta$ , it is evident that classical measurement error is present. Note, however, that  $\text{cov}(\epsilon_j, \epsilon_j) \neq 0$  because the error component  $\hat{\Theta} - \Theta$  is common to all values.

Alternatively, we may combine the tinea likelihood with the likelihood from the phantom studies and maximize them with respect to  $(\mathcal{B}, \Theta)$  jointly. We call this calibrated likelihood. Let  $\ell_{\text{phantom}}(\Theta)$  represent the log-likelihood function from the phantom study. The combined log likelihood is therefore

$$\ell_{\text{combined}}(\mathcal{B}, \Theta) = \ell_{\text{phantom}}(\Theta) + \sum_{j=1}^J \left[ O_j \log \left\{ h(t | D_j^\Theta, X_j, \mathcal{B}) \right\} - h(t | D_j^\Theta, X_j, \mathcal{B}) E_j \right]. \quad (6)$$

This seems awkward in that the states  $j$  are defined for an unknown value of  $\Theta$ . Using a Fisher scoring algorithm for maximization, however, it turns out that the tabulations only need to be formed for current estimates of  $\Theta$  at each iteration.

## 3. Modeling the Radiation Dose Uncertainties

### 3.1 Introduction

To use the likelihood function in (6), we need an expression for  $D^\Theta$ , i.e., a model for expected dose given the observable dose predictor variables and other covariates. Here we describe the available variables for the tinea capitis subjects for predicting their dose and the phantom experiments and physical considerations that are used to translate these predictor variables into a dose determination. We shall give a rough explanation of the process in this introduction and then provide more details and specific models in Sections 3.2 and 3.3.

The phantoms in the phantom experiments were artificial heads and torsos constructed from actual skeletal bone and artificial soft tissue. These were submitted to x-ray exposures using similar machines and a variety of qualitative and quantitative characteristics that matched some of the x-ray prescriptions actually used on the children in Israel.

It is convenient to let  $\mathcal{V}$  represent the vector of dose-predictor variables that would really be needed in order to

make direct use of the phantom study results for predicting dose, and to let  $\mathcal{W}$  represent the dose-predictor variables that are actually available in the existing database. The point of this notational distinction is that  $\mathcal{W}$  contains variables that are incomplete or in some way proxies for  $\mathcal{V}$ .

Specifically, the components of  $\mathcal{V}$  are the number of x-ray applications (between one and four, since some children needed more than one administration due to reoccurrence of ring worm), the ages of the child at each application (between 0 and 16 years), the prescribed beam exposure in roentgens (R) at each exposure, and the added filtration (in mm of aluminum [Al]) used at each exposure. The actually recorded variables ( $\mathcal{W}$ ) for each child are their number of x-ray applications, a code for treatment center or centers at which the x-ray therapies were administered (most applications were at one of three centers), and the age of the child at the time of their first exposure. Also available are the types of x-ray machines that were used at most of the centers and the prescribed exposure and filtration used at most of the centers. Note that the ages at second and subsequent x-ray applications and the order of treatment centers for children who received x-rays at more than one center are unknown.

The individual dose estimates used in the previous analyses of the tinea capitis study group ignored some of the complications. It was thought, based on the phantom studies results, that the thyroid dose for a 6-year-old child receiving a single x-ray administration was about 0.09 Gy (gray). An adjustment was made for younger and older children based on typical distances between the radiation source and the thyroid glands, so that the dose for older children was less than 0.09 and for younger children was greater than 0.09: the specific adjustment is detailed in Section 3.2. This dose was then multiplied by the number of applications to arrive at an estimated total dose. This might tend to overestimate doses for those 9% of the children who received more than one application since it makes no allowance for their increase in age.

Our approach uses more formal statistical modeling. From the relevant phantom studies, we (a) devise and estimate models for  $E(D | X, \mathcal{V}, \Theta)$ ; from considerations of the study population and with some additional assumptions, we (b) devise models for the distribution of  $\mathcal{V}$  given  $X$  and  $\mathcal{W}$ ; and putting these two together gives (c), a model for  $E(D | X, \mathcal{W})$  and therefore the ingredients for the likelihood function in (6). These three tasks are detailed in Sections 3.2, 3.3, and 4, respectively.

### 3.2 Dosimetry from Experiments on Phantoms

Several studies experimentally estimated radiation thyroid doses associated with tinea capitis radiotherapy by exposing anthropomorphic phantoms to similar x-ray conditions (Schulz and Albert, 1968; Werner, Modan, and Davidoff, 1968; Lee and Youmans, 1970; Modan et al., 1977). From the early dosimetry studies, it was believed that conditions like those in Israel would produce absorbed doses to the thyroid of about 0.06 Gy on a 6-year-old child treated a single time. Modan et al. (1977) believed that doses would tend to be larger for live children, who might have been positioned imperfectly and who would have moved during the course of treatment. They investigated the effect of slight repositioning of the phantoms prior to exposure and found a 6-year-old dose to be closer to 0.09 Gy.

Our models for  $E(D | \mathcal{V}, X, \Theta)$  and  $\ell_{\text{phantom}}(\Theta)$  are based on the combined data from the Lee and Youmans (1970) and Modan et al. (1977) studies, with an indicator variable to represent whether movement had been simulated, a presumed physical model for age adjustment, and some additional random effects. The data are included in the  $\ell_{\text{phantom}}(\Theta)$  portion of the combined likelihood in (6).

There is insufficient data from the anthropomorphic phantom studies to adequately estimate the effect of age on dose. Results from similar experiments on water phantoms and mathematical models were used to extrapolate the results to children of other ages. The result may be expressed through multiplicative adjustment factors,  $C_A$ , which specify the dose for a child of age  $A$  relative to the dose for a child of age 6 exposed under the same conditions. The dose adjustment factors for ages 1, 2, ..., 15 are as follows: 1.70, 1.50, 1.39, 1.25, 1.10, 1.00, 0.90, 0.82, 0.74, 0.66, 0.63, 0.60, 0.59, 0.58, and 0.56. For example, all else equal, the dose for a 1-year-old child would be 1.7 times the dose of a 6-year-old.

After considering sources of uncertainty and analyzing the phantom data of Modan et al. (1977) and Lee and Youmans (1970), we developed the following model for the distribution of dose,  $D$ , on a single course of treatment from rounded (to the nearest integer) age of exposure,  $A$ , added filtration (in mm of aluminum),  $F$ , and prescribed beam exposure (in roentgens)  $R$ :

$$\log(D) = \log(C_A R) + \theta_0 + \theta_1 F^2 + \epsilon_w + \epsilon_b + \epsilon_r, \quad (7)$$

where  $\Theta = (\theta_0, \theta_1)$  are unknown parameters and the  $\epsilon$ 's are random error terms representing within-individual effects, between-individual effects, and random errors due to differences between prescribed and actual skin exposure. The error terms are taken to have mean zero and standard deviations  $\sigma_w$ ,  $\sigma_b$ , and  $\sigma_r$ , respectively.

The random error  $\epsilon_w$  represents a within-individual effect, reflecting the different thyroid doses that would occur if a child were hypothetically irradiated twice under identical conditions. The sources of this term are primarily movement during treatment and peculiarities in positioning the body (and shield) for treatment. An estimate of the standard deviation  $\sigma_w$  obtained from the Modan et al. (1977) phantom study, in which a 7-year-old phantom was repositioned between repeated irradiations, is  $\hat{\sigma}_w = 0.17$  based on 13 d.f. However, this estimate involves speculation that the researchers accurately simulated the movement and positioning of a live child with their manipulations of the phantom and is therefore treated with some skepticism. The Modan et al. study does support the assumption that the within-individual errors for log dose are normally distributed.

The random errors  $\epsilon_b$  represent between-individual effects, reflecting the different thyroid doses for different children of identical rounded ages under ideal machine conditions due to differences in head size and shape. An estimate of  $\sigma_b$  from the three distinct phantoms investigated by Lee and Youmans (1970) is  $\hat{\sigma}_b = 0.49$  on 2 d.f. We will assume normality for the between-individual errors, but this assumption is purely speculative. To the level of roughness of the entire analysis though, it seems innocuous.

We have no data for estimating the standard deviation of  $\epsilon_r$ . A study of a single machine by Schulz and Albert (1968) found that the actual skin exposure might differ from the prescribed amount  $R$  by 15% or more. The physicist among us (MS) believes that around 25% is a better estimate for that type of machine during that era. We will explore a range of values  $\sigma_r$  that includes these possibilities.

If normality is assumed for the three random errors in (7), then standard results for the log-normal distribution imply that the mean dose received by child  $i$  on their  $\ell$ th application, as a function of their age ( $A_{i\ell}$ ), exposure ( $R_{i\ell}$ ), and filtration ( $F_{i\ell}$ ) on the  $\ell$ th application, is

$$E(D_{i\ell} | A_{i\ell}, R_{i\ell}, F_{i\ell}) = C_{A_{i\ell}} R_{i\ell} \exp \left\{ \theta_0 + \theta_1 F_{i\ell} + (\sigma_w^2 + \sigma_b^2 + \sigma_r^2) / 2 \right\}.$$

Letting  $D_i$  represent the total dose for irradiated child  $i$ , which is the sum of doses on all their  $L_i$  applications, and letting  $V_i$  represent the ages, exposures, and filtrations at all the child's x-ray applications, we have

$$D_i^\Theta = E(D_i | V_i, \Theta) = \sum_{\ell=1}^{L_i} C_{A_{i\ell}} R_{i\ell} \exp \left\{ \theta_0 + \theta_1 F_{i\ell} + (\sigma_w^2 + \sigma_b^2 + \sigma_r^2) / 2 \right\}. \tag{8}$$

### 3.3 Distribution of Ideal Dose Predictor Variables

In this section, we detail the models used for  $f(V_i | W_i)$ . The important components of this are the distribution of second and subsequent ages of exposure for those children who received multiple applications and the distribution of beam exposure,  $R$ , and machine filtration,  $F$ , given what is known about where the child received their x-ray treatment or treatments. The age at first exposure  $A_{i1}$  is known for all individuals in the study. A histogram of these is displayed in Figure 1. For those individuals who were irradiated more than once, however, it is only known that their ages at second and subsequent exposures were at least 1 year greater than their age at first exposure and less than 16.

As an approximation to  $f(A_{2i} | A_{1i})$ , we used the observed conditional distribution of first ages greater than  $A_{1i} + 1$ , i.e., the part of the histogram in Figure 1 to the right of  $A_{1i} + 1$ . This choice carries the assumption that, e.g., the relative frequencies of ages 7, 8, 9, and so on for the second treatment of a child whose first treatment was at age 6 are the same as the relative frequencies of those values for the age at first treatment. Although it is impossible to check this assumption, it seems reasonable for the level of approximation required and much better than assuming the second age was the same as the first, as was implicitly assumed for the original dose calculations. The densities  $f(A_{3i} | A_{2i})$  and  $f(A_{4i} | A_{3i})$  were similarly defined in terms of the histogram in Figure 1.

The filtration and nominal exposures depend on the machine and location. Table 1 shows what is known about where the irradiation took place, e.g., a subject whose irradiation code is 03 is known to have had one course of treatment at treatment center 1 and one course of treatment at treatment center 2, but it is not known which came first. Table 2 shows the information known about prescribed skin exposure and filtration for the machines used.

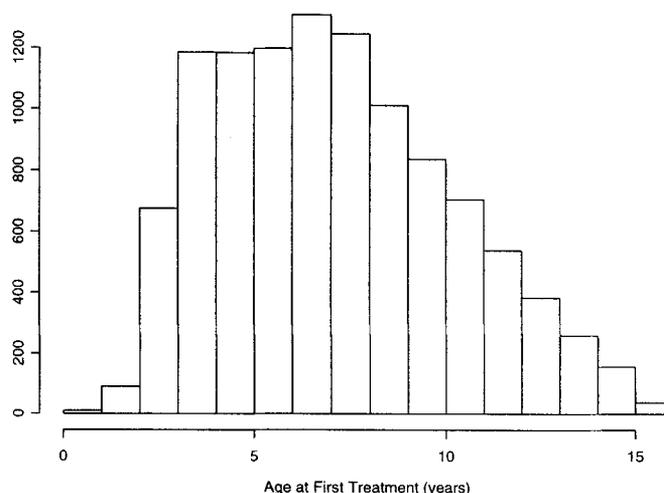


Figure 1. Histogram of the age at first treatment for irradiated subjects.

In treatment center 3, the machines had a common filtration 0.5 mm Al and nominal exposure 425 roentgens. The same occurs in treatment center 2, although the values of the filtration (0.0) and nominal exposure (350) differed from that

Table 1

Description of the places of irradiation and the number of irradiations in the tinea capitis data set. Here  $H$  = treatment center 1;  $T$  = treatment center 2;  $J$  = treatment center 3;  $N$  = unknown;  $O$  = other;  $A$  = abroad. The order of visitation for the centers for those children with courses at several centers has not been retained in the database.

Code for place for irradiation	Number of courses				Centers
	1	2	3	4	
01	7187	394	17	0	H
02	1261	13	0	0	T
03	0	11	1	0	H + T
04	1364	30	1	0	J
05	2	27	4	0	H + J
06	0	3	0	0	T + J
09	0	75	6	0	H + O
10	0	12	0	0	T + O
12	0	4	0	0	J + O
13	0	0	1	0	H + J + O
17	2	140	38	5	H + A
18	0	4	0	0	T + A
20	0	24	8	0	J + A
21	0	0	2	0	H + J + A
33	0	136	18	0	H + N
34	0	3	0	0	T + N
35	0	0	2	0	H + T + N
36	0	28	6	0	J + N
37	0	0	1	0	H + J + N
41	0	0	1	0	H + O + N
49	0	0	3	0	H + A + N
52	0	0	1	0	J + A + N
53	0	0	0	1	H + J + A + N

**Table 2**

Available information about machines used at the various treatment locations; NA means not available. Notice that most treatments were performed in treatment center 1, and only about 3% were performed in *N* (unknown), abroad, or other.

Place	Percentage of all treatments	Machine	Filtration	Prescribed exposure (R)
Treatment center 1	72	1	0.5	400
		2	0.5	384
		3	0.5	383
		4	0.6	NA
Treatment center 3	13	1	0.5	425
		2	0.5	425
		3	0.5	425
Treatment center 2	11	1	0.0	350
		2	0.0	350
N	1.6	1	1.0	350-400
Abroad	1.9		NA	NA
Other	0.8		NA	NA

in treatment center 3. In treatment center 1, there were four machines, with filtrations (0.5, 0.5, 0.5, 0.6) (mm Al) and nominal exposures (400, 384, 383, NA), where NA means unknown, but the machine used on each child was not recorded. Thus, for treatment center 1, we assumed that the actual filtration was 0.5 with probability 0.75 and 0.6 with probability 0.25. For the nominal exposure, we assumed that the distribution was 400, 384, and 383, each with probability 0.25, while, with probability 0.25, the nominal exposure was taken to be uniformly distributed between 350 and 425, reflecting the range of nominal exposures recorded in the various centers. For those recorded as being in site *N* (unknown), the filtration was 1.0 but the nominal exposure was unknown and again taken to be uniformly distributed between 350 and 425. Finally, for those who were irradiated abroad, neither filtration nor nominal exposure were available. The nominal exposure was taken to be uniformly distributed between 350 and 425, while the filtration was assumed to take on the values (0.0, 0.5, 1.0) with probabilities (0.10, 0.85, 0.05), a distribution somewhat in keeping with the observed filtrations. We summarize the distributions of filtration and exposure in Table 3.

Putting the pieces of this section together provides a distributional specification for  $f(\mathcal{V}_i | \mathcal{W}_i)$ . The expectation of total dose for irradiated child  $i$  may be expressed as follows:

$$E(D_i | X_i, \mathcal{W}_i, \Theta) = \sum_{\ell=1}^{L_i} \int E(D_{i\ell} | X_i, \mathcal{V}_i, \Theta) f(\mathcal{V}_i | X_i, \mathcal{W}_i) d\mathcal{V}_i. \quad (9)$$

The covariate vector  $X_i$  has been included in the conditioning statement in (8) to match the notation of the induced hazard function of Section 2. The expressions on the right-hand side of the equation are the same whether or not  $X_i$  is included

**Table 3**

The distribution of filtration  $F_{ij}$  and exposure  $R_{ij}$ . The logarithm of the exposure is assumed to be normally distributed with mean  $\log(\tilde{R}_{ij})$  and standard deviation  $\sigma_r$ . Children exposed in treatment center 1 are assumed to be in one of conditions  $C_1$ - $C_4$  at random, each with probability 0.25. For treatment center 3, the condition is  $C_5$ . For treatment center 2, the condition is  $C_6$ . For location *N* (unknown), the condition is  $C_7$ . For other and abroad, the condition is  $C_8$ .

Condition label	Nominal exposure $\tilde{R}$	Filtration $F$
$C_1$	400	0.5
$C_2$	384	0.5
$C_3$	383	0.5
$C_4$	Uniform[350, 425]	0.6
$C_5$	425	0.5
$C_6$	350	0.0
$C_7$	Uniform[350, 425]	1.0
$C_8$	Uniform[350, 425]	0.0 with probability 0.10 0.5 with probability 0.85 1.0 with probability 0.05

since the important components of  $X_i$  for predicting dose are already included in  $\mathcal{V}_i$  (mainly, age at first exposure).

**4. A Useable Likelihood Function**

The model for the expectation of  $D_i$  given  $\mathcal{V}_i$  and  $X_i$  in (7), along with the expression for  $E(D_i | X_i, \mathcal{W}_i)$  in (9), completes the specification of the combined likelihood function in (6). A Monte Carlo approximation to (9) may be obtained by taking  $M$  samples from the distribution of  $\mathcal{V}_i = (A_{i1}, \dots, A_{iL}, R_{i1}, \dots, R_{iL}, F_{i1}, \dots, F_{iL})$  given  $(X, \mathcal{W})$ . Then approximately

$$E(D_i | X_i, \mathcal{W}_i, \Theta) \approx M^{-1} \sum_{m=1}^M \sum_{\ell=1}^L C_{A_{i\ell m}} R_{i\ell m} \times \exp \left\{ \theta_0 + \theta_1 F_{i\ell m}^2 + \frac{\sigma_w^2 + \sigma_b^2 + \sigma_2^2}{2} \right\}. \quad (10)$$

In the log-likelihood function (5),  $D_j^\ominus$  is the person-years weighted average of the doses for all individuals represented in state  $j$ . Thus, for fixed  $\Theta$ ,

$$D_j^\ominus = M^{-1} \sum_{i \in S_j} \sum_{m=1}^M \sum_{\ell=1}^L w_i C_{A_{i\ell m}} R_{i\ell m} \times \exp \left\{ \theta_0 + \theta_1 F_{i\ell m}^2 + \frac{\sigma_w^2 + \sigma_b^2 + \sigma_2^2}{2} \right\}, \quad (11)$$

where  $S_j$  is the set of indices  $i$  corresponding to individuals whose exposure history includes state  $j$  and  $w_i$  is the person-

**Table 4**

Results for the dose-response model  $\exp(X_{br}^T \beta_{br}) \{1 + D \exp(\beta_{dr} + X_{em}^T \beta_{em})\}$ . Parameter estimates are given, along with standard errors (in parentheses). There is obviously little difference between regression calibration and calibrated likelihood. The results are for within-person standard deviation  $\sigma_w = 0.17$ , between-person standard deviation  $\sigma_b = 0.49$ , and the standard deviation of random errors due to differences between prescribed and actual skin exposure  $\sigma_r = 0.15$ .

	Model with age at exposure as modifying effect of dose		
	Naive	Regression calibration	Calibrated likelihood
Terms background rate			
Constant	-11.94 (.65)	-11.95 (.65)	-11.94 (.65)
Female indicator	1.37 (.33)	1.37 (.33)	1.37 (.33)
Africa indicator	-0.29 (.33)	-0.30 (.32)	-0.29 (.32)
Israel indicator	-0.84 (.46)	-0.84 (.46)	-0.85 (.46)
Attained age in [15, 20)	1.05 (.56)	1.05 (.56)	1.04 (.56)
Attained age in [20, 25)	1.23 (.52)	1.23 (.52)	1.23 (.52)
Attained age in [25, 30)	1.32 (.51)	1.32 (.51)	1.32 (.51)
Attained age in [30, 35)	1.30 (.51)	1.30 (.52)	1.29 (.52)
Attained age [35, $\infty$ )	1.34 (.58)	1.34 (.58)	1.33 (.58)
Dose	-0.38 (.57)	-0.53 (.57)	-0.52 (.58)
Effect modifier			
age at exposure	-0.12 (.09)	-0.11 (.09)	-0.11 (.09)
Maximized log likelihood	-598.0	-597.8	-598.4
Maximized log likelihood without modifying effects	-598.9	-598.6	-599.3
Dose without modifying effects	-1.05 (.41)	-1.18 (.40)	-1.17 (.43)
Maximized log likelihood without dose effects	-614.3	-614.3	-614.9

years of observation of individual  $i$  in state  $j$  as a proportion of the total person-years of observation in state  $j$ . The Fisher scoring algorithm may be used to find the values of  $\mathcal{B}$  and  $\Theta$  that maximize (6) with  $D_j^\Theta$  expressed this way. We treated the variance parameters,  $\sigma_w^2$ ,  $\sigma_b^2$ , and  $\sigma_r^2$ , as known and conducted the analysis with several combinations of values within their speculated ranges.

We may now describe the two approaches for analysis more succinctly. For regression calibration, the value of  $\Theta$  was estimated in a first stage and then treated as known in the second. The expected doses given available dose-predictor variables,  $D_j^\Theta$ , were calculated for each individual using (10). These values were then used in place of true dose in a standard routine for failure time regression via the subject-years approach. The calibrated likelihood approach used the Fisher scoring algorithm to simultaneously estimate  $\mathcal{B}$  and  $\Theta$  in the combined likelihood function (6). The result is approximate likelihood analysis, with the approximation due to the dropping of the condition  $T > t$  in the expected hazard function (3) and due to assuming that covariates take on constant values within states formed by the cross-classification. The cross-classification for sets  $S_j$  in (11), in both cases, was based on sex (two levels), origin (3 levels: Africa, Asia, Israel), age at first exposure (eight levels: [0, 2), [2, 4), [4, 6), [6, 8), [8, 10), [10, 12), [12, 14), [14, 16)), attained age (eight levels: [0, 15), [15, 20), [20, 25), [25, 30), [30, 35), [35, 40), [40, 45), [45+)),

and expected dose (six levels: 0, (0, 7.5), [7.5, 15), [15, 22.5), [22.5, 30), [30, 100)). Attained age at time of thyroid cancer is the response here. Since the patients were all exposed as children, there is little difference between using this response and time since exposure. In the calibrated likelihood approach, the cross-classification is performed at each iteration after an updating of the estimates of  $\Theta$  based on the combined data sets.

## 5. Results

As seen in Table 4, the difference in the estimates from the two approaches is very small relative to the standard errors. Notice, though, that the standard error for the coefficient of

**Table 5**  
The estimated relative risk when children  
are exposed at a dose of 0.1 Gy; 95%  
confidence intervals are given in parentheses.

Model	Age = 1	Age = 6	Age = 15
Naive	7.1 (3.2, 17.6)	4.4 (2.4, 8.8)	2.2 (1.2, 9.7)
Corrected	6.3 (2.9, 15.3)	4.0 (2.2, 8.2)	2.1 (1.1, 9.6)

dose is larger in the calibrated likelihood estimate, as would be expected since this estimate correctly incorporates the uncertainty in the estimate of  $\Theta$ . Neither approach, however, incorporates the uncertainty in the estimates of the variances of the random effects.

The results may be summarized as follows:

1. There is a statistically significant dose-response (one-sided  $p$ -value = .009). Although the parameter estimate for the modifying effect of age on the dose-response increased by 40% when accounting for dosimetry uncertainties, it is not statistically significant (two-sided  $p$ -value = .37). Some specific estimates of relative risk are shown in Table 5.
2. Accounting for error in dosimetry changes hardly any of the parameter estimates or the relative risk for different ages at first exposure.
3. These results are for  $(\sigma_w, \sigma_b, \sigma_r) = (0.17, 0.49, 0.15)$ . We have redone the analysis with  $(\sigma_w, \sigma_b, \sigma_r) = (0.17, 0.25, 0.25)$  and  $(0.5, 0.5, 0.5)$  (as an extreme case), and there is little change in the results. The estimated effect of dose tends to be smaller when larger variances for the random effects are assumed, but the magnitude of the change is quite small relative to its standard error.

## 6. Discussion and Conclusions

Our reanalysis of the effect of thyroid radiation dose on age-specific thyroid cancer rates from the Israeli children irradiated for tinea capitis and their matched controls showed little change from the previous analyses. We shall provide a few comments in conclusion about the general statistical points that may apply to similar data problems with measurement errors.

At the heart of the statistical framework here is the availability of a large primary data set with health outcomes and imprecisely determined doses and a secondary data set with information for predicting dose from available dose-predictor variables. A Berkson error arises from the use of the expected dose (from a regression model) in place of true dose. A classical error arises from the use of estimated coefficients in the regression model. An important point is that the classical errors in dose estimates for different individuals are correlated.

Another epidemiological problem with this structure is the study of health effects of indoor air pollutants, as discussed by Tosteson, Stefanski, and Schafer (1989). Although interest was in a dose-response model for total exposure to nitrogen dioxide, the actual exposures could only be predicted from readings on stationary monitors and prediction equations developed from secondary data sets. Those authors treated the predictions as having Berkson errors, ignoring the classical errors involved in the estimation of the regression equation. In our approach for the tinea capitis data, we have formally combined the prediction problem and the dose-response problem by performing a single likelihood analysis.

In the tinea data problem, the classical errors due to estimating unknowns in the prediction equation turned out not to be consequential. This is apparently because the part of the dose prediction model involving unknown parameters (the intercept and the part having to do with the effect of machine filtration on dose) was not as important as the parts that were free of unknown parameters (the parts having to do with the

effects of beam exposure and age). Since the model of interest here was linear in dose, the Berkson error had little effect and was easy to handle. In nonlinear-in-dose models, the effect of the Berkson error might be consequential and linearization may be needed to approximate the induced hazard function given available predictor variables, corresponding to (4).

Much of the discussion about the general statistical structure for the tinea study is overshadowed by the high degree of speculation involved in transporting the dose prediction model from the phantoms to the live children. We introduced random effects for additional sources of uncertainty and explored results for various presumed values for their standard deviations. While there are likely to be inaccuracies in modeling and missed sources of uncertainty, the lack of sensitivity of the results to changes in the standard deviations offers some assurance that the uncertainties may be handled without more precise modeling.

## ACKNOWLEDGEMENTS

Carroll's research was supported by a grant from the National Cancer Institute (CA57030) and through the Texas A&M Center for Environment and Rural Health via a grant from the National Institute of Environmental Health Sciences (P30-ES09106). Stovall's research was supported by a contract from the National Cancer Institute (N01-CP-91024). We thank Professor Baruch Modan for permission to use the tinea capitis data and an anonymous referee for helpful comments on organization.

## RÉSUMÉ

En Israël durant les années 40 et 50, plus de 20 000 enfants ont été traités pour le tinea capitis (teigne) par irradiation pour provoquer l'épilation. Les études de suivi ont montré que la dose de radiations était associée au développement de tumeurs malignes de la thyroïde. Malgré la nette évidence d'un tel effet, l'ampleur de la relation dose-réponse est moins claire à cause des erreurs sur l'estimation de la dose reçue par la glande thyroïdienne. Ces erreurs peuvent induire dans l'estimation de l'effet de dose un biais, lequel n'a pas été totalement apprécié dans les études initiales. On revient sur ce fait en décrivant d'une façon détaillée comment les erreurs de dosimétrie peuvent se produire et en développant un nouveau modèle "dose-réponse" qui prend en compte les incertitudes sur la dosimétrie. Ce modèle est une nouvelle variante complexe du modèle multifactoriel classique de l'erreur de Berkson avec des composantes sur les erreurs de mesure et sur les données manquantes. L'analyse des données sur le tinea capitis suggère que l'erreur de mesure sur la dosimétrie a un effet négligeable sur l'estimation de l'effet dose-réponse, sur l'inférence statistique et sur l'effet modificateur de l'âge au moment de l'exposition.

## REFERENCES

- Breslow, N. E., Lubin, J. H., Marek, P., and Langholz, B. (1983). Multiplicative models and cohort analysis. *Journal of the American Statistical Association* **78**, 1-12.
- Carroll, R. J., Rupert, D., and Stefanski, L. A. (1995). *Measurement Error in Nonlinear Models*. London: Chapman and Hall.
- Cox, D. R. and Oakes, D. (1984). *Analysis of Survival Data*. London: Chapman and Hall.

- Hu, P., Tsiatis, A. A., and Davidian, M. (1998). Estimating the parameters in the Cox model when covariate variables are measured with error. *Biometrics* **54**, 1407–1409.
- Hughes, M. D. (1993). Regression dilution in the proportional hazards model. *Biometrics* **49**, 1056–1066.
- Lee, W. and Youmans, H. D. (1970). *Doses to the central nervous system of children resulting from x-ray therapy for tinea capitis*. Technical Report BRH/DBE. U.S. Department of Health, Education, and Welfare, Public Health Service, Washington, D.C.
- Lubin, J. H., Boice, J. D., and Samet, J. M. (1995). Errors in exposure assessment, statistical power and the interpretation of residential radon studies. *Radiation Research* **144**, 329–341.
- Modan, B., Ron, E., and Werner, A. (1977). Thyroid cancer following scalp irradiation. *Radiology* **123**, 741–744.
- Nakamura, T. (1992). Proportional hazards models with covariates subject to measurement error. *Biometrics* **48**, 829–838.
- Pepe, M. S., Self, S. G., and Prentice, R. L. (1989). Further results in covariate measurement errors in cohort studies with time to response data. *Statistics in Medicine* **8**, 1167–1178.
- Pierce, D. A., Stram, D. O., and Vaeth, M. (1990). Allowing for random errors in radiation dose estimates for the atomic bomb survivors. *Radiation Research* **123**, 275–284.
- Pierce, D. A., Stram, D. O., Vaeth, M., and Schafer, D. (1992). Some insights into the errors in variables problem provided by consideration of radiation dose–response analyses for the A-bomb survivors. *Journal of the American Statistical Association* **87**, 351–359.
- Prentice, R. L. (1982). Covariate measurement error and parameter estimation in a failure time regression model. *Biometrika* **69**, 331–342.
- Ron, E. and Hoffman, F. O. (1999). *Uncertainties in Radiation Dosimetry and Their Impact on Dose response Analysis*. Bethesda, Maryland: National Cancer Institute Press.
- Ron, E., Modan, B., Preston, D., Alfandary, E., Stovall, M., and Boice, J. D. (1989). Thyroid neoplasia following low-dose radiation in childhood. *Radiation Research* **120**, 516–531.
- Ron, E., Lubin, J. H., Shore, R. E., Mabuchi, K., Modan, B., Pottern, L. M., Schneider, A. B., Tucker, M. A., and Boice, J. D. (1995). Thyroid cancer after exposure to external radiation: A pooled analysis of seven studies. *Radiation Research* **141**, 259–277.
- Schulz, R. J. and Albert, R. E. (1968). Follow-up study of patients treated by x-ray epilation for tinea capitis III: Dose to organs of the head from the x-ray treatment of tinea capitis. *Archives of Environmental Health* **17**, 935–950.
- Thomas, D., Stram, D., and Dwyer, J. (1993). Exposure measurement error: Influence on exposure–disease relationships and methods of correction. *Annual Review of Public Health* **14**, 69–93.
- Tosteson, T., Stefanski, L. A., and Schafer, D. W. (1989). A measurement error model for binary and ordinal regression. *Statistics in Medicine* **8**, 1139–1147.
- Werner, A., Modan, B., and Davidoff, D. (1968). Doses to the brain, skull and thyroid following x-ray therapy for tinea capitis. *Physics in Medicine and Biology* **13**, 247–258.

Received September 2000. Revised April 2001.

Accepted April 2001.

## APPENDIX

### Derivation of (3)

As in Prentice (1982) and Pepe et al. (1989), the induced hazard function is

$$\begin{aligned}
 h(t | X, \mathcal{W}, \mathcal{B}) &= \lim_{\Delta \rightarrow 0} \frac{\text{pr}(t \leq T \leq t + \Delta | X, \mathcal{W}, \mathcal{B})}{\Delta} \\
 &= \lim_{\Delta \rightarrow 0} \int \frac{\text{pr}(t \leq T \leq t + \Delta | D, X, \mathcal{W}, \mathcal{B}, T \geq t)}{\Delta} \\
 &\quad \times f(D | X, \mathcal{W}, T \geq t, \Theta) dD \\
 &= \int \lim_{\Delta \rightarrow 0} \frac{\text{pr}(t \leq T \leq t + \Delta | D, X, \mathcal{B}, T \geq t)}{\Delta} \\
 &\quad \times f(D | X, \mathcal{W}, T \geq t, \Theta) dD,
 \end{aligned}$$

this last step following from the conditional independence assumption (2). We have thus shown, as claimed, that  $h(t | X, \mathcal{W}, \mathcal{B}) = E\{h(t | D, X, \mathcal{B}) | X, \mathcal{W}, T \geq t, \Theta\}$ .