

## **The Use of Joint Models in Analysis of Aggregate Endpoints in RERF Cohort Studies**

Authors: Sposto, Richard, and Cullings, Harry M.

Source: Radiation Research, 201(4) : 304-309

Published By: Radiation Research Society

URL: <https://doi.org/10.1667/RADE-23-00122.1>

---

BioOne Complete ([complete.BioOne.org](https://complete.BioOne.org)) is a full-text database of 200 subscribed and open-access titles in the biological, ecological, and environmental sciences published by nonprofit societies, associations, museums, institutions, and presses.

Your use of this PDF, the BioOne Complete website, and all posted and associated content indicates your acceptance of BioOne's Terms of Use, available at [www.bioone.org/terms-of-use](https://www.bioone.org/terms-of-use).

Usage of BioOne Complete content is strictly limited to personal, educational, and non - commercial use. Commercial inquiries or rights and permissions requests should be directed to the individual publisher as copyright holder.

---

BioOne sees sustainable scholarly publishing as an inherently collaborative enterprise connecting authors, nonprofit publishers, academic institutions, research libraries, and research funders in the common goal of maximizing access to critical research.

# The Use of Joint Models in Analysis of Aggregate Endpoints in RERF Cohort Studies

Richard Sposto,<sup>a,1</sup> Harry M. Cullings<sup>b</sup>

<sup>a</sup> Department of Statistics, and <sup>b</sup> Expert Advisor, Department of Statistics, Radiation Effects Research Foundation, 5-2 Hijiyama Park, Minami Ku, Hiroshima City, 732-0815, Japan

---

Sposto R, Cullings HM. The Use of Joint Models in Analysis of Aggregate Endpoints in RERF Cohort Studies. *Radiat Res.* 201, 304–309 (2024).

In radiation risk estimation based on the Radiation Effects Research Foundation (RERF) cohort studies, one common analysis is Poisson regression on radiation dose and background and effect modifying variables of an aggregate endpoint such as all solid cancer incidence or all non-cancer mortality. As currently performed, these analyses require selection of a surrogate radiation organ dose, (e.g., colon dose), which could conceptually be problematic since the aggregate endpoint comprises events arising from a variety of organs. We use maximum likelihood theory to compare inference from the usual aggregate endpoint analysis to analyses based on joint analysis. These two approaches are also compared in a re-analysis of RERF Life Span Study all cancer mortality. We show that, except for a trivial difference, these two analytic approaches yield identical inference with respect to radiation dose response and background and effect modification when based on a single surrogate organ radiation dose. When repeating the analysis with organ-specific doses, an interesting issue of bias in intercept parameters arises when dose estimates are undefined for one sex when sex-specific outcomes are included in the aggregate endpoint, but a simple correction will avoid this issue. Lastly, while the joint analysis formulation allows use of organ-specific doses, the interpretation of such an analysis for inference regarding an aggregate endpoint can be problematic. To the extent that analysis of radiation risk for an aggregate endpoint is of interest, the joint-analysis formulation with a single surrogate dose is an appropriate analytic approach, whereas joint analysis with organ-specific doses may only be interpretable if endpoints are considered separately for estimating dose response. However, for neither approach is inference about dose response well defined. © 2024 by Radiation Research Society

---

## INTRODUCTION

The Life Span Study (LSS) of the Radiation Effects Research Foundation (RERF) is used to estimate the added

<sup>1</sup> Corresponding author address: Department of Statistics, Radiation Effects Research Foundation, 5-2 Hijiyama Park, Minami Ku, Hiroshima City, 732-0815, Japan; email: rsposto@rerf.or.jp.

risk of cancer and non-cancer incidence or mortality due to exposure to ionizing radiation (1–5). Analyses of LSS data can include those based on aggregate disease endpoints, such as all solid cancer incidence, all solid cancer mortality, or all non-cancer mortality, usually supplemented by analyses of selected organ-specific endpoints. As currently performed, the analysis of these aggregate endpoints involves selection of a surrogate, representative dose, typically colon dose [e.g., (2)]. Conceptually this might be viewed as unsatisfactory since the individual endpoints that comprise the aggregate endpoint can arise from organs different from the surrogate organ dose (e.g., occurrence of breast cancer or occurrence of thyroid cancer, arising from organs anatomically distinct from the colon). Also, it is not clear that colon is the organ most representative of the entirety of organs, or of organs that seem to be susceptible to radiation-induced cancer, as it is one of the deepest organs in the body and therefore has the lowest ratio of neutron to gamma-ray dose (6).

To the extent that analyses of radiation risk on aggregate disease endpoints are of interest, it is possible to formulate these analyses instead as joint analysis in which events arising from different organs are treated individually (7). In this paper, we compare the usual formulation with an alternative joint formulation of the analysis of aggregate endpoints, first using analytic methods, specifically maximum likelihood theory, and then in a reanalysis of all solid cancer mortality data from Report 14 of the RERF Life Span Study (4). We also discuss the implications of using a representative dose versus organ specific doses in such analyses.

## METHODS

### *Analytic Comparison of the Usual Formulation and the Alternative Joint Formulation*

Analyses of RERF cancer incidence, cancer mortality, and non-cancer mortality have typically been based on Poisson regression methods (8, 9) using a summary of the data into a table comprising a large number of cells cross-classified into categories of variables such as age at exposure, radiation dose, sex, etc., with, within each cell, accumulated person-years of exposure and numbers of events, as well as categorical values or person-year-weighted means of variables that are used as covariates in analysis.

Let

- $i$  reference cells in the person-year table,  $i = 1, \dots, l$ ;
- $P_i$  be the number of person-years accumulated in cell  $i$ ;
- $j$  reference a specific mode of failure (e.g., stomach cancer mortality)  $j = 1, \dots, J$ ;
- $y_{ij}$  be the number of cases of mode  $j$  in cell  $i$ ;
- $X_i$  be a covariate vector of length  $p$  associated with cell  $i$ ;
- $\beta_j$  be the parameter vector of length  $p$  for mode  $j$  for covariate vector  $X_i$ ;
- $\alpha_j$  be the scalar intercept parameter for failure mode  $j$ .

Note that  $P_i$  is the same for all modes of failure  $j$ , as follow-up is censored for occurrence of any of the modes of failure.  $X_i$ , which can include person-year weighted means of continuous variables, will also be the same for all failure modes.

Letting the rate function  $\lambda(X_i, \alpha_j, \beta_j) = \omega(\alpha_j)\theta(X_i, \beta_j)$ , the most general Poisson likelihood is

$$\prod_i \prod_j \frac{[P_i \omega(\alpha_j) \theta(X_i, \beta_j)]^{y_{ij}} e^{-P_i \omega(\alpha_j) \theta(X_i, \beta_j)}}{y_{ij}!}$$

In the case of an aggregate endpoint, we are interested in fitting a model that is agnostic to the failure mode  $j$ , in this case  $\lambda(X_i, \alpha, \beta) = \omega(\alpha)\theta(X_i, \beta)$  for scalar parameter  $\alpha$  and  $p$ -vector parameter  $\beta$ . Letting  $\omega(\alpha) = \omega$  and  $\theta(X_i, \beta) = \theta_i$  to simplify notation, the usual formulation of this problem uses the Poisson likelihood based on  $y_i = \sum_j y_{ij}$ , with log-likelihood

$$l_U = \sum_i \{y_i \ln(P_i \omega \theta_i) - P_i \omega \theta_i - \ln(y_i!)\} \tag{1}$$

whereas an alternative joint formulation (assuming  $\alpha_j = \alpha$  and  $\beta_j = \beta$  for all  $j$ ) is based on  $y_{ij}$  with log-likelihood

$$\begin{aligned} l_A &= \sum_i \sum_j \{y_{ij} \ln(P_i \omega \theta_i) - P_i \omega \theta_i - \ln(y_{ij}!)\} \\ &= \sum_i \{y_i \ln(P_i \omega \theta_i) - J P_i \omega \theta_i\} - \sum_i \sum_j \ln(y_{ij}!) \end{aligned} \tag{2}$$

where subscripts  $U$  and  $A$  will denote ‘‘usual’’ and ‘‘alternative.’’ We assume that  $\omega(\cdot)$  and  $\theta(\cdot, \cdot)$  are twice differentiable and monotone functions with respect to  $\alpha$  and the elements of  $\beta$ . Let  $\omega'$  and  $\omega''$  be the first and second derivatives of  $\omega$  with respect to  $\alpha$ , and  $\theta'_i$  be the first partial derivative  $1 \times p$  matrix and  $\theta''_i$  the second partial derivative  $p \times p$  matrix of  $\theta_i$  with respect to the elements of  $\beta$ . The score equations with respect to parameters  $\alpha$  and  $\beta$  are:

$$\begin{aligned} l_U^{(\alpha)} &= \sum_i \left\{ y_i \frac{\omega'}{\omega} - P_i \omega' \theta_i \right\} \\ l_U^{(\beta)} &= \sum_i \left\{ y_i \frac{\theta'_i}{\theta_i} - P_i \omega \theta'_i \right\} \end{aligned}$$

and

$$\begin{aligned} l_A^{(\alpha)} &= \sum_i \left\{ y_i \frac{\omega'}{\omega} - J P_i \omega' \theta_i \right\} \\ l_A^{(\beta)} &= \sum_i \left\{ y_i \frac{\theta'_i}{\theta_i} - J P_i \omega \theta'_i \right\}. \end{aligned}$$

Note that  $l_U^{(\beta)}$  and  $l_A^{(\beta)}$  represent  $p$  equations each, one for each partial derivative with respect to the elements of  $\beta$ . Denoting the solutions to  $l_U^{(\alpha)} = 0$  and  $l_U^{(\beta)} = 0$  with hat ( $\hat{\cdot}$ ) and the solutions to  $l_A^{(\alpha)} = 0$  and  $l_A^{(\beta)} = 0$  with tilde ( $\tilde{\cdot}$ ), then

$$\hat{\omega} = \frac{y_{\cdot\cdot}}{\sum_i P_i \hat{\theta}_i}$$

and

$$\tilde{\omega} = \frac{y_{\cdot\cdot}}{J \sum_i P_i \tilde{\theta}_i}$$

Substituting these quantities into the corresponding ( $\beta$ ) score equations yields

$$l_U^{(\beta)} = \sum_i \left\{ y_i \frac{\hat{\theta}'_i}{\hat{\theta}_i} - y_{\cdot\cdot} \frac{P_i \hat{\theta}'_i}{\sum_r P_r \hat{\theta}_r} \right\}$$

and

$$l_A^{(\beta)} = \sum_i \left\{ y_i \frac{\tilde{\theta}'_i}{\tilde{\theta}_i} - y_{\cdot\cdot} \frac{P_i \tilde{\theta}'_i}{\sum_r P_r \tilde{\theta}_r} \right\}.$$

As these last two equations are identical, it follows that  $\hat{\beta} = \tilde{\beta}$ , i.e., the maximum likelihood estimates of  $\beta$  are identical in these two formulations. Furthermore,  $\hat{\omega} = J \tilde{\omega}$ .

It is shown in the appendix that in the special case where  $\omega(\alpha) = e^\alpha$ , (i.e.,  $\ln$  link), the estimated variance-covariance matrix is identical in the two formulations, so that precision estimates of the maximum likelihood estimates of  $\beta$  are identical in the two formulations.

#### Use of Organ-Specific Doses

An interesting issue arises when one considers the use of organ-specific doses when sex-specific endpoints are included in analysis, since organ doses for incongruous sex/site combinations (such as uterine dose for males or prostate doses for females) will be undefined, and as a result the corresponding person-year table cells, which can contain no positive event counts in any case, would be excluded in a joint analysis of an aggregate endpoint. We initially thought that this indicated that the person-years for these sex/site combinations should also be excluded from an analysis even using a single surrogate dose, reasoning that males cannot accumulate person-years at risk for female-specific cancers and vice versa. However, this is not correct, as it is a fact that men are not at risk for female cancers and women are not at risk for male cancers, so that these person-years of follow-up should be included when discussing aggregate endpoints such as all solid cancer incidence. By excluding these cells, the included data are no longer completely representative of the population of interest and this can lead to bias in estimates of sex-specific background parameters in the sense that the resulting estimates are no longer consistent for the same values as in the standard analysis of the aggregate endpoint. In order to make a fair comparison between a joint analysis using surrogate dose with a joint analysis using organ specific doses, it is necessary to correct for the exclusion of the cells with incongruous sex/site combinations.

In Eq. (2) above we assumed implicitly that all modes of failures ( $j$ ) had the same number of contributing cells ( $i$ ). Alternatively, let  $I_{ij}$  be a 0/1 indicator of whether a cell  $i$  is included for failure mode  $j$ , where cells with  $I_{ij} = 0$  are those that contribute person years but cannot contribute cases (i.e.,  $y_{ij} = 0$ ), such as those from incongruous sex/site combinations with undefined dose estimates. Equation (2) can be rewritten as

$$\begin{aligned} l_A &= \sum_i \sum_j \{y_{ij} \ln(P_i \omega \theta_i) - P_i \omega \theta_i - \ln(y_{ij}!)\} \\ &= \sum_i \sum_j I_{ij} \{y_{ij} \ln(P_i \omega \theta_i) - P_i \omega \theta_i - \ln(y_{ij}!)\} \\ &\quad - \sum_i (J - I_{\cdot\cdot}) P_i \omega \theta_i \end{aligned}$$

$$= \sum_i \{y_i \ln(P_i \omega \theta_i) - I_i P_i \omega \theta_i\} - \sum_i \sum_j \ln(y_{ij}!) - \sum_i (J - I_i) P_i \omega \theta_i$$

since by definition  $I_{ij} y_{ij} = y_{ij}$  and  $I_{ij} \ln(y_{ij}!) = \ln(y_{ij}!)$ .  $I_i$  is the number of modes of failure included for cell  $i$ . The first two terms on the right-hand side (RHS) of line 3 comprise the objective function that is maximized if the cells with  $I_{ij} = 0$  are excluded. This is different from Equation 2 and therefore can yield estimates that differ compared to the standard analysis. However, if we re-weight the person-years  $P_i$  by a correction factor  $J/I_i$ , then, the first term on the RHS becomes

$$I_A^* = \sum_i \left\{ y_i \ln \left( \frac{J}{I_i} P_i \omega \theta_i \right) - I_i \frac{J}{I_i} P_i \omega \theta_i \right\} - \sum_i \sum_j \ln(y_{ij}!) = \sum_i \{y_i \ln(P_i \omega \theta_i) - J P_i \omega \theta_i\} + \sum_i y_i \ln \left( \frac{J}{I_i} \right) - \sum_i \sum_j \ln(y_{ij}!)$$

This objective function will yield the same score equations as Eq. (2) above, and hence the same maximum likelihood estimates.

*Poisson Regression Reanalysis of all Solid Cancer Mortality*

RERF Life Span Study Report 14 (4) all solid cancer mortality data were reanalyzed to illustrate the differences and similarities between these analytic formulations.

The source data comprise a 53,782 cell person-year table summarizing outcome of 86,661 LSS subjects stratified by city (Hiroshima, Nagasaki), sex (male, female), ground distance from hypocenter (<3 km, 3–10 km), Adult Health Study (AHS) participation (not AHS, AHS original cohort, AHS additional cohort), age at exposure (5-year categories), attained age (5-year categories), calendar time (5-year categories), DS02 colon dose with RBE = 10 weighting (22 categories).

The total of 10,929 solid cancer deaths comprise deaths from each of 17 individual causes of solid cancer death: esophagus, stomach, colon, rectum, liver, gallbladder, pancreas, other digestive, lung, breast, uterus, ovary, prostate, bladder, kidney, other urinary, and other solid cancer. Organ-specific radiation doses were also included for each of these cancer sites as shown in Table 1. As the current RERF dosimetry system only provides dose estimates for 15 distinct organs (10), a surrogate dose was assigned for some organs, as per Ozasa et al. (4). A complete description of this study cohort can be found in Ozasa et al. (4).

Note that uterine, ovary, and prostate cancers are sex specific, occurring in either males or females only. Breast cancer, while predominantly occurring in women, was not considered sex specific as it does rarely occur in men, breast dose is defined for men, and deaths from male breast cancer were reported in the LSS. Testicular cancer should be treated as sex-specific, but these were included in the “other solid cancer” category in the Ozasa et al. (4) data and could not be separated out in this analysis.

For joint analysis, the person-year table was replicated for each of the 17 causes of death, creating a 914,294-cell table with an additional factor indicating the organ-specific cause (7).

A linear excess relative risk Poisson regression model of the form

$$\lambda(X_0, D, X_1) = e^{\alpha + X_0 \beta_0} [1 + \beta_D D e^{X_1 \beta_1}]$$

was fit, with dependent variable the number of solid cancer deaths, scalar parameters  $\alpha$  and  $\beta_D$ , and vector parameters  $\beta_0$  and  $\beta_1$ .

The independent variables were

$X_0$  comprising:

- city (Nagasaki),

TABLE 1

DS02 Organ Doses Assigned to Specific Organ Sites

Organ site	DS02 Organ Dose
Esophagus*	Stomach
Stomach	Stomach
Colon	Colon
Rectum*	Bladder
Liver	Liver
Gall bladder*	Liver
Pancreas	Pancreas
Other digestive system*	Colon
Lung	Lung
Breast	Breast
Uterus	Uterus
Ovary	Ovary
Prostate*	Bladder
Bladder	Bladder
Kidney*	Colon
Other urinary*	Colon
Other solid tumor*	Colon

Note. \*Denotes organ sites for which a surrogate organ dose was used (please see text).

- sex (Female),
- $\ln(\text{attained age (years)}/70)$ ,
- $\ln(\text{attained age}/70)^2$ ,
- $\ln(\text{attained age}/70)^2$  for attained age > 70,
- age at exposure (years) – 30,

D: RBE 10 weighted DS02 radiation dose in Gy,

$X_1$  comprising:

- age at exposure – 30,
- $\ln(\text{attained age}/70)$ ,
- sex (Female).

Note that this model differs from that used in Ozasa et al. table 4 (4) in that in the background is not parametrically modelled but rather stratified. In addition, the sex effect modifier was a multiplicative factor  $1 + \sigma \cdot s$  where  $s = 1$  for males and  $s = -1$  for females, whereas we use an indicator for female sex. The parametric background model we used was similar to that shown by Grant et al. (2). Our analysis results are essentially equivalent but not identical to those of Ozasa et al. (4).

Five different analyses were performed:

1. Dependent variable total solid cancer deaths, colon dose [usual formulation].
2. Dependent variable organ specific solid cancer deaths, colon dose [alternative formulation].
3. Dependent variable organ specific solid cancer deaths, colon dose, cells corresponding to incongruent sex/site combinations are excluded (i.e., mortality from uterine and ovarian cancer for males, and mortality from prostate cancer for females), without the person-year correction factor described above. [alternative formulation].
4. Same as 3, but with the person-year correction factor described above. [alternative formulation].
5. Same as 4, but using organ specific dose rather than colon dose [alternative formulation].

Excess relative risk models were fit using the gnm package in R (11) using R functions written for this purpose (12).

**TABLE 2**  
**Parameter Estimates and Standard Errors from Five Different Analyses**

Parameter	Analysis 1		Analysis 2/4		Analysis 3		Analysis 5	
	Estimate	(S.E)	Estimate	(S.E)	Estimate	(S.E)	Estimate	(S.E)
Intercept	4.59	(0.0189)	1.76	(0.0189)	1.89	(0.0189)	1.76	(0.0189)
Intercept (adjusted)*	–	–	4.59	(0.0189)	4.72	(0.0189)	4.59	(0.0189)
X <sub>0</sub> : city (Nagasaki)	0.0234	(0.0210)	0.0234	(0.0210)	0.0234	(0.0210)	0.0243	(0.0210)
X <sub>0</sub> : sex (Female)	–0.680	(0.0212)	–0.680	(0.0212)	–0.744	(0.0212)	–0.681	(0.0213)
X <sub>0</sub> : log(attained age/70)	4.57	(0.125)	4.57	(0.125)	4.57	(0.125)	4.57	(0.125)
X <sub>0</sub> : log(attained age/70) <sup>2</sup>	–0.386	(0.229)	–0.386	(0.229)	–0.386	(0.229)	–0.372	(0.228)
X <sub>0</sub> : log(attained age/70) <sup>2</sup> age > 70	–2.37	(0.936)	–2.37	(0.936)	–2.37	(0.936)	–2.38	(0.935)
X <sub>0</sub> : age at exposure - 30	0.0185	(0.00734)	0.0185	(0.00734)	0.0185	(0.00734)	0.0191	(0.00735)
D: dose (Gy)	0.232	(0.0454)	0.232	(0.0454)	0.232	(0.0454)	0.239	(0.0445)
X <sub>1</sub> : age at exposure - 30	–0.0929	(0.0676)	–0.0929	(0.0676)	–0.0929	(0.0676)	–0.104	(0.0654)
X <sub>1</sub> : log (attained age/70)	–1.97	(0.391)	–1.97	(0.391)	–1.97	(0.391)	–1.86	(0.382)
X <sub>1</sub> : sex (Female)	0.928	(0.204)	0.928	(0.204)	0.928	(0.204)	0.921	(0.196)

Analysis 1 = Usual analysis using colon dose. Analysis 2 = Joint analysis using colon dose. Analysis 3 = Joint analysis using colon dose, incongruent sex/site cells excluded, person-years are not adjusted with correction factor. Analysis 4 = Joint analysis using colon dose, incongruent sex/site cells excluded, person-years are adjusted using correction factor. Analysis 5 = Joint analysis using organ-specific dose, incongruent sex/site cells excluded, person-years are adjusted using correction factor. \*Intercept estimate adjusted by adding  $\ln(\# \text{ sites}) = \ln(17)$  (please see text). (Analysis 2 and Analysis 4 results are combined, as they are numerically identical.)

## RESULTS

The parameter estimates and standard errors from these five analyses are shown in Table 2. Analysis 1 is the usual analysis based on solid cancer deaths summed over all types, as per Eq. (1), and Analysis 2 is the alternative joint analysis corresponding to Eq. (2). Note that, except for the intercept parameter, the parameter estimates and their standard errors are identical in the two analyses. The intercept parameter is the  $\alpha$  in  $\omega(\alpha) = e^\alpha$ . Since the number of different cancer sites is 17, per the previous section,  $e^{4.59} = 17 \times e^{1.76}$ , so adding  $\ln(17)$  to the intercept estimate of Analysis 2 gives the intercept estimate of the usual analysis, Analysis 1, and since adding a constant has no effect on the standard error, adjusting the Analysis 2 intercept estimate in this way makes Analysis 1 and Analysis 2 identical.

In comparing our analysis to that of Ozasa et al. (4) table 4, because of the different parametrization of the sex effect modifier, our ERR estimate of 0.232, which refers to males, should be multiplied by  $\frac{1}{2}(1 + e^{0.928})$  to yield a so-called sex-averaged estimate of 0.4094, similar to 0.42 reported in Ozasa et al. (4).

In Analysis 3, the cells corresponding to incongruent sex/site combinations are dropped – 25,068 cells each for ovarian and uterine cancer death in males, and 28,714 cells for prostate cancer death in females. Comparing Analysis 3 to Analysis 2, while other parameter estimates are not affected by this change, the intercept and sex background modifier parameters are affected, with a 14% difference in the former and 7% difference in the latter on the natural scale.

Analysis 4 applies the person-year correction factor and yields estimates that are numerically identical to those in Analysis 2. Analysis 5 differs from Analysis 4 only in that organ doses specific to the cancer type (Table 1) are used

rather than the surrogate colon dose. Here dose response and a number of other parameter estimates differ, but only slightly compared to Analyses 2 and 4.

## DISCUSSION

In this paper, we prove analytically and demonstrate by example that a joint analysis formulation of an aggregate disease endpoint will yield identical inference to the usual approach when using a single surrogate representative dose. Use of joint analysis for an aggregate endpoint has negligible impact on analytic complexity, requiring only an expanded person-year table which could also be used for site specific analyses that usually are also performed.

Our analysis revealed an issue of possible bias in estimates of sex-specific background parameters that results if sex-specific outcomes are included but cells with incongruous sex/site combinations are excluded. In our example, other parameters, notably radiation dose response parameters, were unaffected. Whether and to what extent this would be the case with other model parameterizations that may include interactions between sex and other variables is unclear. The bias can be corrected with a simple person-year correction factor.

Joint analysis has been advocated in the past as a way coherently to analyze multiple endpoints (7) for the reasons that such analyses allow sharing of parameters in models of different cancer types, allowing formal significance tests for differences in parameters among cancer types, the ability to formulate unified models, and the use of organ-appropriate radiation dose estimates. This last could be especially relevant when RERF adopts the new organ dosimetry that is currently being developed, as this will provide specific doses for a much larger collection of organ sites (13), each having much different ratios of neutron to gamma-ray doses (6).

Broadly speaking, the concept of dose response for an aggregate endpoint using a surrogate dose is ill defined. There is no exposure in the real world that is uniform to all tissues, and all exposures are different. Even in a constant radiation field for external exposure, for example, organ doses depend on the person's size, orientation, age, sex, and other physiological features. When discussing the increased risk to an individual of, say, solid cancer mortality from 1 Gy colon dose exposure, there is an implicit but unstated assumption that the relationship between colon dose and doses to other organs is approximately constant in ratio for all individuals. This means, though, that in applying models based on RERF cohorts to aggregate endpoint risk in other cohorts based on the surrogate dose, one is again assuming implicitly that the same relationship between other organ doses and the surrogate dose obtains. In this case the surrogate dose simply serves as a useful index without any clear biological relevance to the aggregate outcome.

Joint analysis of an aggregate endpoint using organ specific doses, as also discussed in this paper, is even less well defined. On the one hand it is not clear to what the dose term in the prediction model refers. On the other hand, this analysis is simply the most parsimonious joint model that can be fit to such data, and in this sense is admissible, although it is very unlikely that a final joint model would share all background, dose response, and effect modifier parameters among endpoints so that the model could be interpreted as relevant to any of the individual endpoints. The use of organ specific doses in joint analysis may really only makes sense when dose response parameters are estimated separately for individual disease sites or disease groups that share the same organ dose.

The event rate function for an aggregate endpoint derives from a combination of endpoint-specific population rate functions, each of which is based on organ specific dose, and which, even if having the same underlying functional form, will have different background, dose response, and/or effect modifier parameters. Recognizing this underlying heterogeneity, there are other ways to construct a risk model for an aggregate endpoint. For example, one could fit individual and independent parametric models of the rate function for each specific endpoint, or alternatively fit a joint model with some shared and some endpoint-specific parameters, as suggested by Pierce and Preston (7). The resulting endpoint-specific rate functions could be summed to obtain the rate function applicable to the aggregate endpoint. If each endpoint-specific model is an accurate description of the failure process for that endpoint, then the resulting predictions from the summed failure rate models are coherent in that they derive directly from those of these individual models. Even though prediction might still be indexed by a surrogate organ dose, the method would allow utilization of different implicit ratios of each organ dose to the surrogate dose when applying RERF results to populations where these ratios may differ significantly. These approaches are worthy of further study.

## APPENDIX

For the usual formulation, since

$$\hat{\omega} = \frac{y_{..}}{\sum_r P_r \hat{\theta}_r}$$

the estimated information matrix components are

$$l_U^{(\hat{\alpha})}(\hat{\alpha}) = \sum_i \left\{ y_i \frac{\hat{\omega}'' \hat{\omega} - \hat{\omega}' \hat{\omega}'}{\hat{\omega}^2} - P_i \hat{\omega}'' \hat{\theta}_i \right\}$$

$$= - \frac{\hat{\omega}' \hat{\omega}'}{y_{..}} \left( \sum_r P_r \hat{\theta}_r \right)^2$$

$$l_U^{(\hat{\beta})}(\hat{\beta}) = - \sum_i P_i \hat{\omega}' \hat{\theta}_i'$$

$$l_U^{(\hat{\beta})}(\hat{\beta}) = - \sum_i \left\{ y_i \frac{\hat{\theta}_i'' \hat{\theta}_i - \hat{\theta}_i'^T \hat{\theta}_i'}{\hat{\theta}_i^2} - P_i \hat{\omega} \hat{\theta}_i'' \right\}$$

$$= - \sum_i \left\{ y_i \frac{\hat{\theta}_i'' \hat{\theta}_i - \hat{\theta}_i'^T \hat{\theta}_i'}{\hat{\theta}_i^2} - y_{..} \frac{P_i \hat{\theta}_i''}{\sum_r P_r \hat{\theta}_r} \right\}$$

Similarly, for the alternative formulation, since

$$\tilde{\omega} = \frac{y_{..}}{J \sum_r P_r \tilde{\theta}_r}$$

the estimated information matrix components are

$$l_A^{(\tilde{\alpha})}(\tilde{\alpha}) = \sum_i \left\{ y_i \frac{\tilde{\omega}'' \tilde{\omega} - \tilde{\omega}' \tilde{\omega}'}{\tilde{\omega}} - J P_i \tilde{\omega}'' \tilde{\theta}_i \right\}$$

$$= - J^2 \frac{\tilde{\omega}' \tilde{\omega}'}{y_{..}} \left( \sum_r P_r \tilde{\theta}_r \right)^2$$

$$l_A^{(\tilde{\beta})}(\tilde{\beta}) = - J \sum_i P_i \tilde{\omega}' \tilde{\theta}_i'$$

$$l_A^{(\tilde{\beta})}(\tilde{\beta}) = - \sum_i \left\{ y_i \frac{\tilde{\theta}_i'' \tilde{\theta}_i - \tilde{\theta}_i'^T \tilde{\theta}_i'}{\tilde{\theta}_i^2} - J P_i \tilde{\omega} \tilde{\theta}_i'' \right\}$$

$$= - \sum_i \left\{ y_i \frac{\tilde{\theta}_i'' \tilde{\theta}_i - \tilde{\theta}_i'^T \tilde{\theta}_i'}{\tilde{\theta}_i^2} - y_{..} \frac{P_i \tilde{\theta}_i''}{\sum_r P_r \tilde{\theta}_r} \right\}$$

In the special case where  $\omega(\alpha) = e^\alpha$ , i.e., *ln* link, then  $\omega = \omega' = \omega''$ , and

$$l_U^{(\hat{\alpha})}(\hat{\alpha}) = l_U^{(\hat{\beta})}(\hat{\beta}) = l_A^{(\tilde{\alpha})}(\tilde{\alpha}) = l_A^{(\tilde{\beta})}(\tilde{\beta}) = - y_{..},$$

and since  $\hat{\beta} = \tilde{\beta}$ , the entire estimated information matrix, as well as its inverse, the estimated variance-covariance matrix, are identical in the two formulations.

### ACKNOWLEDGMENTS

The Radiation Effects Research Foundation (RERF), Hiroshima and Nagasaki, Japan is a public interest foundation funded by the Japanese Ministry of Health, Labour and Welfare (MHLW) and the US Department of Energy (DOE). The research was also funded in part through DOE award DE-HS0000031 to the National Academy of Sciences. This publication was supported by RERF Research Protocols 1-75. The views of the authors do not necessarily reflect those of the two governments.

Received: June 21, 2023; accepted: January 25, 2024; published online: February 13, 2024

### REFERENCES

1. Brenner AV, Preston DL, Sakata R, et al. Comparison of all solid cancer mortality and incidence dose-response in the Life Span Study of atomic bomb survivors, 1958-2009. *Radiat Res* 2022; 197(5):491-508.
2. Grant EJ, Brenner A, Sugiyama H, et al. Solid Cancer Incidence among the Life Span Study of Atomic Bomb Survivors: 1958-2009. *Radiat Res* 2017; 187(5):513-37. doi: 10.1667/RR14492.1
3. Ozasa K, Cullings HM, Ohishi W, et al. Epidemiological studies of atomic bomb radiation at the Radiation Effects Research Foundation. *Int J Radiat Biol* 2019; 95(7):879-91. doi: 10.1080/09553002.2019.1569778
4. Ozasa K, Shimizu Y, Suyama A, et al. Studies of the mortality of atomic bomb survivors, Report 14, 1950-2003: an overview of cancer and noncancer diseases. *Radiat Res* 2012; 177(3):229-43. doi: 10.1667/RR2629.1
5. Preston DL, Shimizu Y, Pierce DA, et al. Studies of mortality of atomic bomb survivors. Report 13: Solid cancer and noncancer disease mortality: 1950-1997. *Radiat Res* 2003; 160(4):381-407. doi: RR3049
6. Cordova KA, Cullings HM. Assessing the relative biological effectiveness of neutrons across organs of varying depth among the atomic bomb survivors. *Radiat Res* 2019; 192(4):380-7.
7. Pierce DA, Preston DL. Joint analysis of site-specific cancer risks for the atomic bomb survivors. *Radiat Res* 1993; 134(2):134-42.
8. Frome EL. The Analysis of Rates Using Poisson Regression Models. *Biometrics* 1983; 39(3):665-74.
9. Frome EL, Morris MD. Evaluating Goodness of Fit of Poisson Regression Models in Cohort Studies. *The American Statistician* 1989; 43(3):144-47. doi: 10.1080/00031305.1989.10475642
10. Cullings HM, Fujita S, Funamoto S, et al. Dose estimation for atomic bomb survivor studies: its evolution and present status. *Radiat Res* 2006; 166(1 Pt 2):219-54. doi: 10.1667/RR3546.1
11. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing 2022.
12. RERF Department of Statistics. R-Library for Excess Relative Risk (ERR) and Excess Absolute Risk (EAR) Models 2022 [Available from: [https://www.rerf.or.jp/en/about/organization-en/chart-e/statistics/analysistools\\_e/](https://www.rerf.or.jp/en/about/organization-en/chart-e/statistics/analysistools_e/)].
13. Griffin KT, Sato T, Funamoto S, et al. Japanese pediatric and adult atomic bomb survivor dosimetry: Potential improvements using the J45 phantom series and modern Monte Carlo transport. *Radiat Environ Biophys* 2021; 1-14.