Allowance for dose-estimation errors in RERF data

Donald A. Pierce
Retired RERF

Primary reference for this talk is Pierce, Vaeth, Cologne, Allowance for Random Dose-Estimation Errors in Atomic Bomb Survivor Studies; A Revision. *Rad. Res.* 2008, p. 118

This can be obtained by connecting to
www.science.oregonstate.edu/~piercedo/

Other useful references are given in that paper

After far too much struggling, we decided to replace *estimated dose* by approximation to *E(true | estimated) ,* along with related things such as replacing *true^2* by *E(true^2 | estimated)*

Literature in 1986 served us poorly, but also we struggled with Bayesian vs frequentist considerations --- is a given survivor's true dose a random variable? (see our JASA paper, also Gilbert 1984)

Necessarily involves *pr(true, estimated)* , which for *measurement-type error model* (classical) is represented as *pr(true) pr(estimated | true) .*

Methods not explicitly addressing *pr(true)* for the cohort are suspect. Best to set aside "zero" doses for this modeling. Problematic to use validation data to assess *E(true | estimated)* since distribution of true doses there usually differs from that for the cohort

We largely rely more heavily than some on external considerations and sensitivity analysis for *pr(estimated | true) .* Everything here assumes this is lognormal, with attention largely restricted to choice of CV for this

Our original methods for this were rough and cumbersome --- involving crude parametric assumptions, numerical deconvolution and numerical integration

Only much later were able to deal with *averaging-type* (Berkson-type) errors, in manner similar to others (e.g. RCDW). Clearer definition of Berkson-type errors would seem useful

Not an emphasis here, but much later found that lognormal for *pr(true | nonzero)* would have been reasonably adequate. Reeves at al (RCDW) an important paper

Then, $\log E(\text{true} \mid \text{est}) = R^2 \log(\text{est}) + (1 - R^2) \log E(\text{true})$, with

$$R^2 = \text{var(true)} / \{ \text{var(true)} + \text{var(est} \mid \text{true)} \}$$

In 2004 Kellerer and I developed improved derivations that (see PVC):

(a) allow for both *measurement-type* and *averaging-type* (classical and Berkson-type) errors, and

(b) provided explicit expression for E(*true* | *estimated*) in terms of log variances of the two types of error, along with aspects of pr(*true*)

Regarding (a) we assumed, similarly to others, that

$$\log(estimated) = \log(true) + e\_meas + e\_avg$$

Where *e_meas* is independent of *true* and *e_avg* is independent of *estimated* , each with specified variances

Result for (b) , based on Laplace approximation generalizing lognormal results, is as follows. There are similar results for E(*true^2* | *estimated*) for use in LQ models and other needs

Suppose one knew the empirical distribution of true doses for the cohort. Write $d_1(z), d_2(z)$ for the first two log-log derivatives of the density for this, evaluated at any given estimated dose z

Then to excellent approximation, writing

$$E(true \mid z) = C(z)\ z$$

we have that

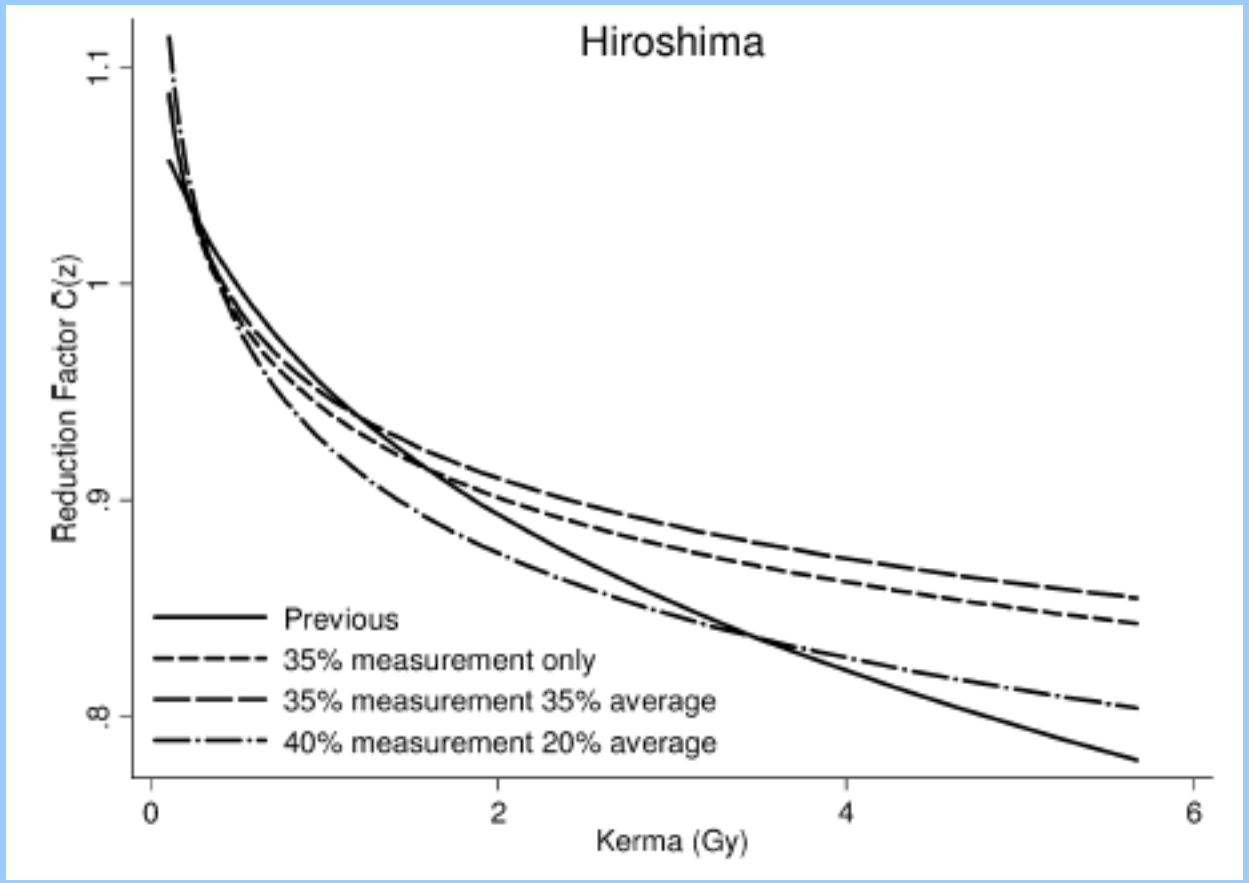
$$\log[C(z)] = \frac{1 + 2d_1(z)}{1 - \sigma_M^2 d_2(z)} \frac{\sigma_M^2}{2},$$

For the lognormal case $d_1(z) = \{E(x) - z)\} / \mathrm{var}(x)$, $d_2(z) = -1/\mathrm{var}(x)$ yielding result on previous slide

Now write $\hat{d}_1(z), \hat{d}_2(z)$ for the log-log derivatives based on the distribution of *estimated* doses, then

$$d_1(z) = \hat{d}_1(z)/[1 + (\sigma_M^2 - \sigma_A^2)\hat{d}_2(z)]$$
$$d_2(z) = \hat{d}_2(z)/[1 + (\sigma_M^2 - \sigma_A^2)\hat{d}_2(z)].$$

Reduction factors *C(z)* for Hiroshima, under various assumptions.



Most telling results of all this come from resulting ease of sensitivity analysis:

TABLE 1
Results of Various Adjustments as Shown in Fig. 1 on Risk Estimation for Solid Cancer Mortality

Method	ERR (30)	Increase	Exposure age	City (P)	Curvature (P)
Unadjusted	0.42		−0.38	1.14 (0.14)	0.56 (0.02)
Previous adjustment	0.47	1.12	−0.38	1.18 (0.09)	1.10 (0.007)
35% measurement (new)	0.47	1.12	−0.38	1.16 (0.12)	1.11 (0.007)
35% measurement 35% average	0.46	1.10	−0.38	1.17 (0.11)	1.04 (0.008)
40% measurement 20% average	0.48	1.14	−0.38	1.17 (0.11)	1.28 (0.005)

Some adjustment is useful, although with modest effect, but reasonable variation in assumptions has nearly negligible effect. Averaging-type (Berkson) errors actually reduce adjustments

I think that parameter values specific to major shielding categories may be counterproductive, as could be other unwarranted focus on precision

Our recommendation for present is as in the last line of this table

Further uncertainties best dealt with by considering

$$y = \beta' true + \beta' \{true - E(true \mid est)\} + error$$

where middle term is random overdispersion

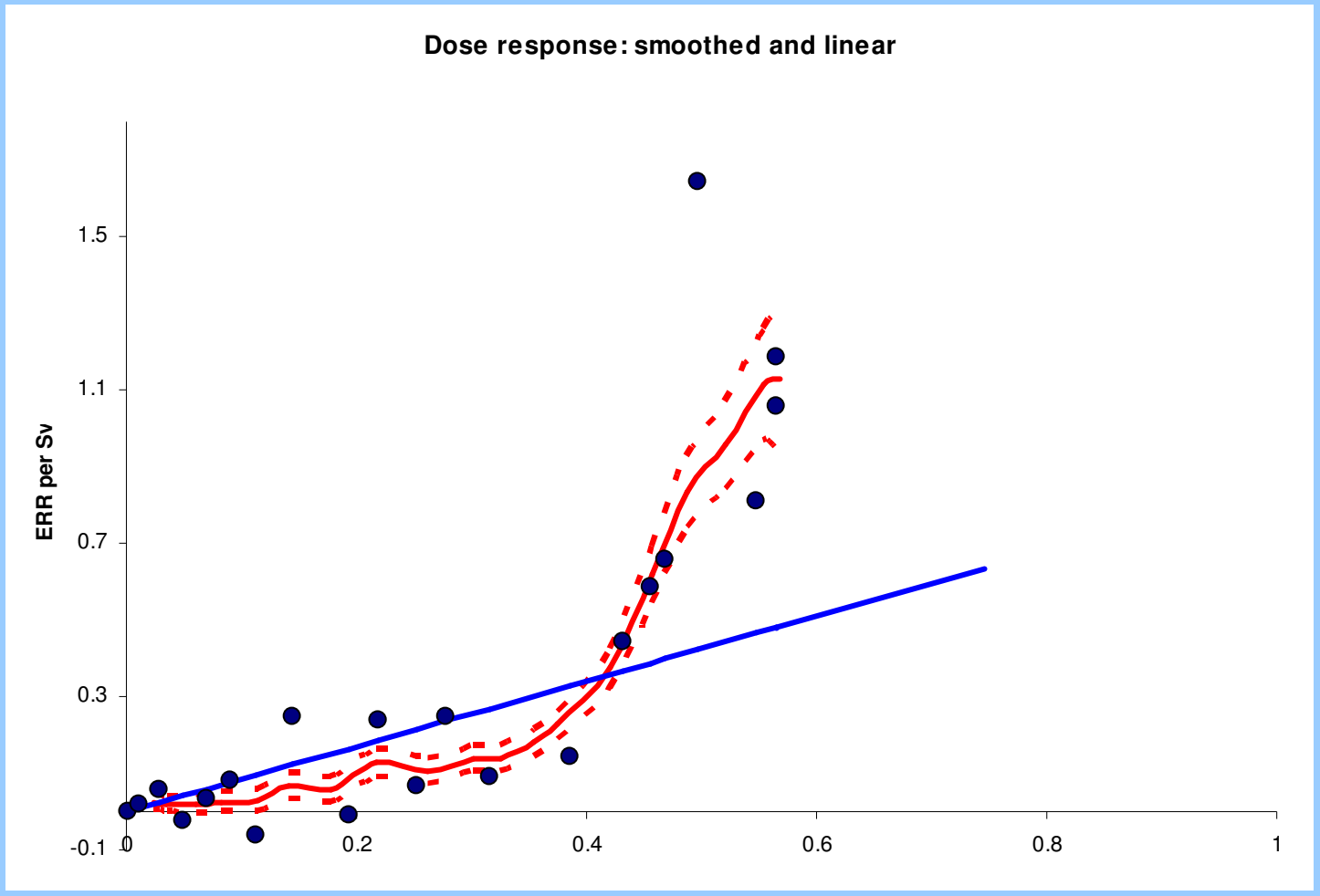
For binary/surival data, *var(error)* usually dominates this overdispersion, but for other binomial data it may not, and overdispersion must be dealt with

In the latter case, methods at RERF for chromosome data have long attended to the overdispersion term (even before understanding it)

Remarkable consequence is that in former case, estimation of parameter is negligibly degraded by dose-estimation errors of reasonable magnitude

However, further work on this, and effects on standard errors, will be useful (underway, with slow progress on my part)

Effects of assuming much larger errors



Shape of (adjusted) cancer dose-response assuming 100% CV dose-estimation errors. The linear fit shown has slope about twice the usual estimates, but no one would use linear risk estimates in view of these results

Assessment of the derivative functions

$$\hat{d}_1(z)\ ,\hat{d}_2(z)$$

from distribution of estimated doses