Computing RERF Survivor Dose Estimates

Donald A Pierce

April 2016 for Stat-Epi Depts

Preface

- I plan to update this by sometime in early May, to be clearer about the matter of Measurement and Averaging errors, and also about the simulation results
- We will post another message when this is available

In 1986-87 it took Stram, Vaeth and me extensive effort to understand that the key issue in dose-estimation errors pertains to

avg(true dose | estimated dose)

where this average is over the analysis cohort. Approximations to these funtions of estimated dose are at RERF called “adjusted dose estimates” or “survivor dose estimates”.

Will next indicate why --- even when the dose estimates are unbiased --- we will have that Avg(true | estimated) < estimated.

Adjusting for this type of bias is the essence of adjustments to deal with dose estimation error. We are currently working on implementing a method for this that was developed in 2004-2008.

These issues have to do with utilizing the information provided by the survival of the subject, which it not intended to be used in DS86 – DS02

Artificial Example Indicating Basic Concepts

True dose x (Gy)	Estimated dose z (Gy) 1	2	3	4	5	6	Number of survivors
—	—						—
—	—	—					—
2	250	500	250				1000
3		75	150	75			300
4			33	66	33		132
5				15	30	15	60
—					—	—	—
Avg($x \mid z$)	—	—	2.50	3.62	—	—	

The rapid decline in numbers of survivors with increasing true dose is largely a “survival” matter. This decline is reflected in each column of estimated dose.

Note that for this example, among those with estimated dose 3 Gy the average true dose is here 2.50 --- the adjusted dose estimate

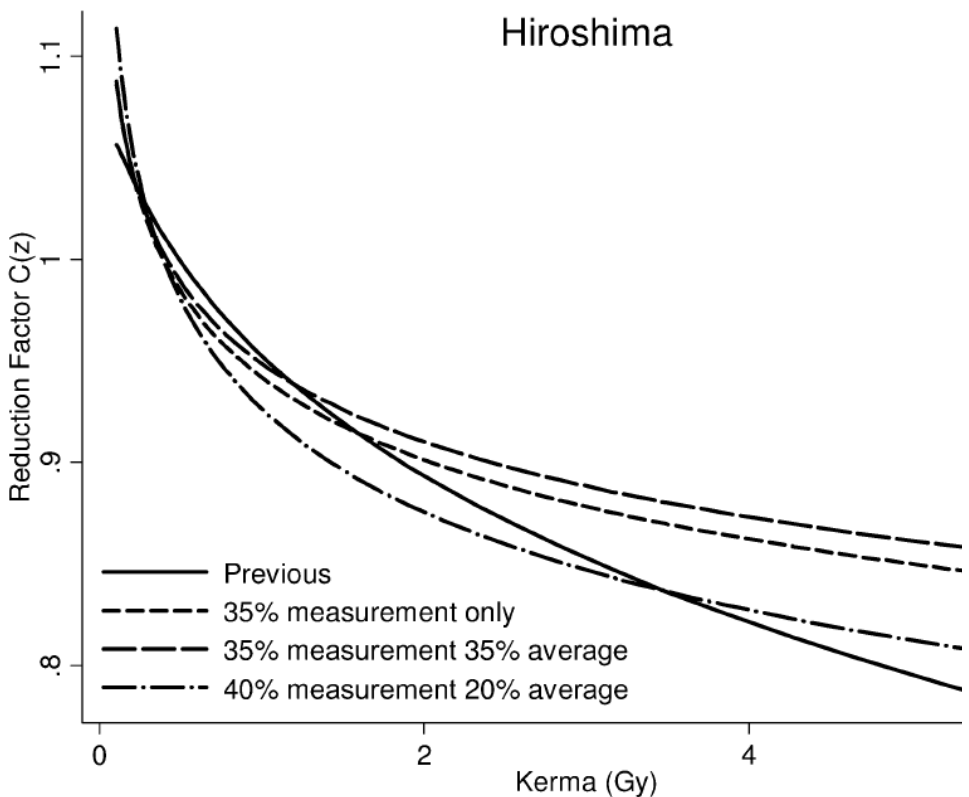
The method used for current adjustments was far too cumbersome for present needs.

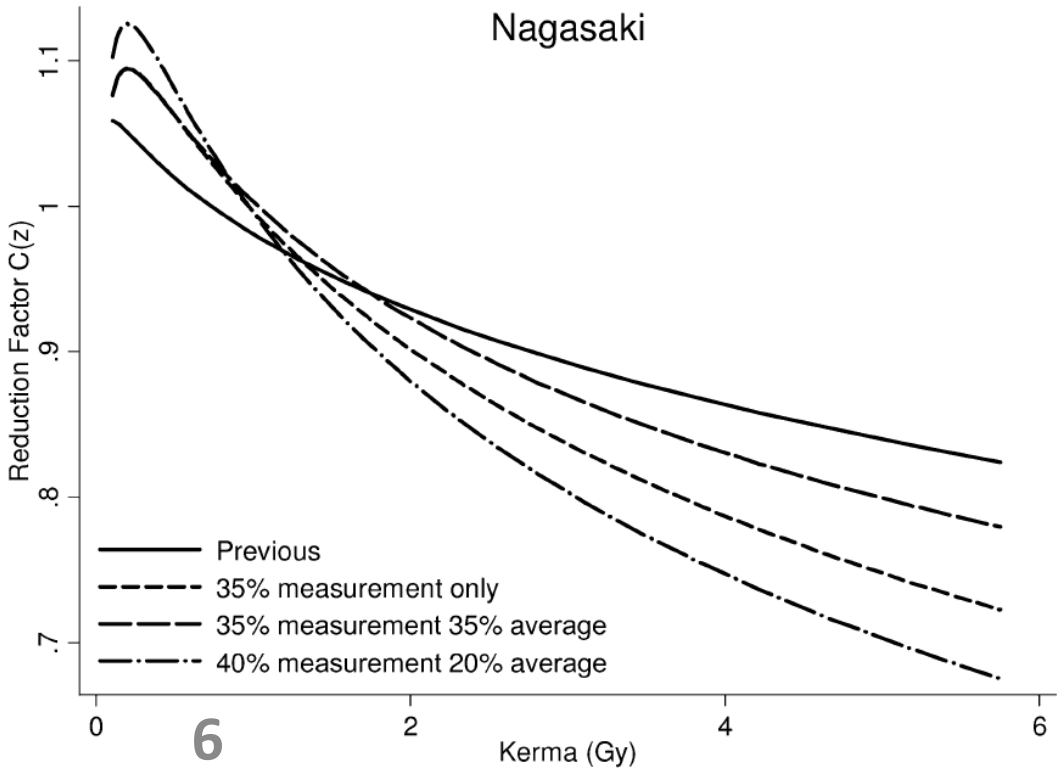
Did not allow for the two types of dose-estimation errors: Measurement (the classical form) and Grouping (Berkson-type).

Could not readily be automated for use of differing assumed magnitudes of errors. Important in evaluating the performance of the methods employed in reducing the risk-estimation (attenuation) bias due to dose-estimation errors.

Advances were made during 1994-2008 that overcome many of these limitations, and we propose implementing them now as the standard method.

Using the new formula, and developments to allow for both measurement and grouping error, it is easy now to carry out explorations such as this. These are dose-adjustment factors for various assumptions on the magnitude and form of dose-estimation errors.





Symbolically, writing x for true dose and z for estimated dose, we have that the joint distribution of these, and the adjusted dose estimates, are

$$p(z,x) = p(x)\ p(z|x)$$

$$p(x|z) = c(z)\ p(x)\ p(z|x)$$

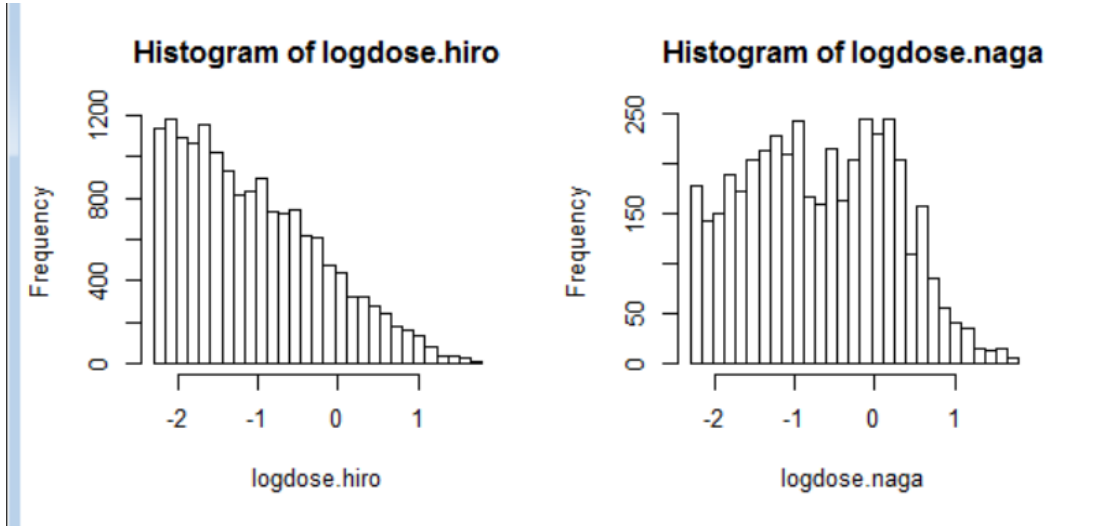
$$Avg(x|z) = Sum\{x\ p(x|z)\}$$

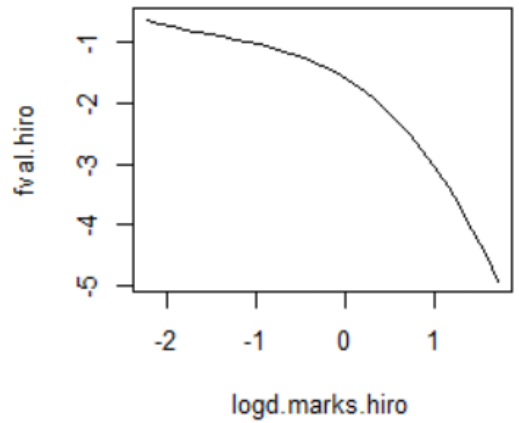
Here $p(x)$ is the cohort distribution of true dose and $p(z|x)$ represents the model for dose-estimation errors (typically lognormal). Methods here are documented in the 2008 RadRes paper

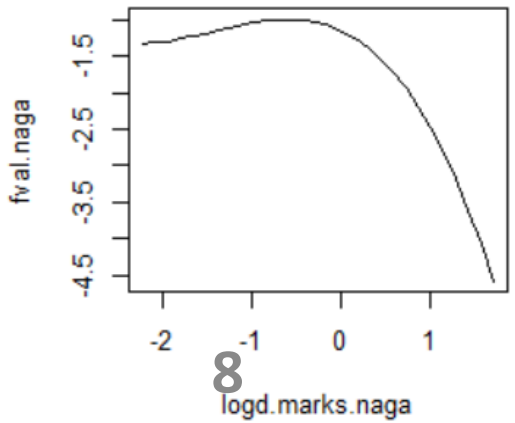
Allowance for Random Dose Estimation Errors in Atomic Bomb Survivor Studies: A Revision

Donald A. Pierce,[a,1] Michael Væth[b] and John B. Cologne[c]

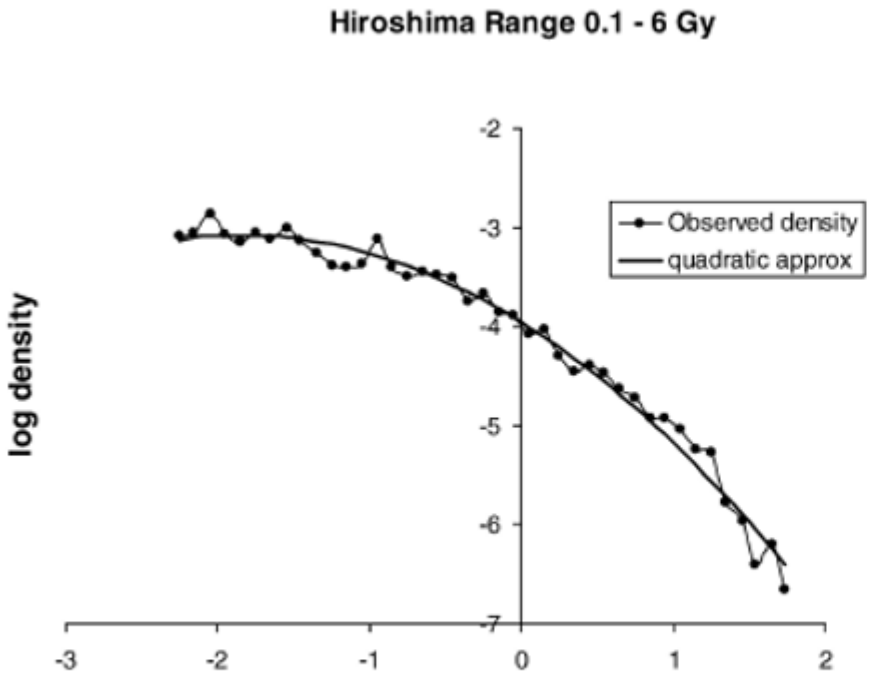
Moving on, consider, on log-log scale, the distribution of estimated doses and approximating cubic function for describing those

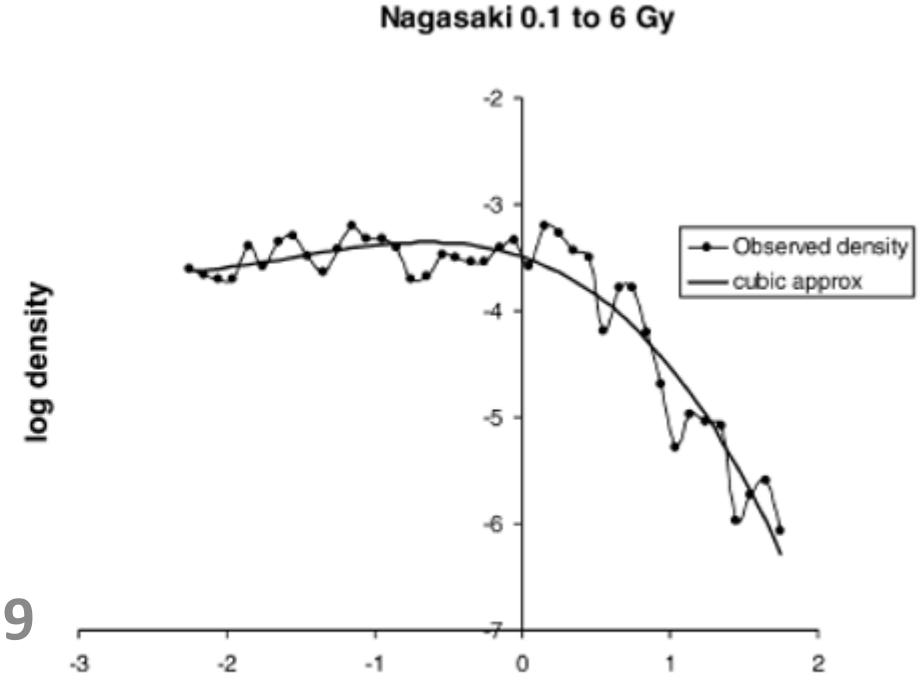






This shows more clearly the adequacy of those type of approximations, for DS02 but very near results for DS86. Here the Hiro curve is taken as quadratic rather than cubic. Interestingly this is a lognormal model, fitting better (for positive doses) than we had thought. Lognormal *p(x)* means that log(x) is normally distributed





Those curves were developed by exploratory analysis, so this is nonparametric. In practiced now we use cubic curves in place of this exploratory analysis. An issue discussed later is that these must be done in terms of estimated, not true, dose.

The adjustment factor C(*z)* to be applied to DS02 dose estimates is given by

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

where *d1(z) is the slope of curves as just shown, evaluated at z,* and *d2(z)* is the rate of change of that slope. Here σ_M is the SD of log measurement error, which is approximately the SD of *z* relative to *x*.

This formula arose as follows. For the case that both *p(x)* and *p(z|x)* are lognormal, it is well known that

log {*E(x|z)*} = *w* log*(x)* + *(1-w)* log(*z*)

Where

w = SD{ log(*z*) } I x } / SD { log(*x*) }

with approximation SD{ log(*z*) | *x* } = CV(*z* | *x*) = SD(*z* | *x*) / *x*

Further, in the relations here for given *z* , and other related comsoderations,

$$p(z,x) = p(x)\ p(z|x)$$

$$p(x|z) = c(z)\ p(x)\ p(z|x)$$

$$Avg(x|z) = Sum\{x\ p(x|z)\}$$

we only need to consider variations in *x* that are moderately near to *z* .

Thus it suffices to assume *p(x)* is locally lognormal, which is broadly true

This “local” approach was suggested by Kellerer, a classically trained physicist, where the method is attributed to Laplace. This Laplace method is important in modern methods to improve on classical likelihood inference, which is my other main research interest. In the relations

$$p(z,x) = p(x)\ \ p(z|x)$$
$$p(x|z) = c(z)\ \ p(x)\ \ p(z|x)$$
$$Avg(x|z) = Sum\{x\ p(x|z)\}$$

this is used to approximate the final line by taking *p(x|z)* as locally quadratic, which corresponds to taking *p(x)* as locally lognormal, for which the classical results at the top of the previous slide apply exactly. The weights *w* there can be calculated from the local approximation, in terms of the functions *d1(z)* and *d2(z)*. When the lognormal approximation to *p(x)* is global rather than only local, the function *d2(z)* is identically zero.

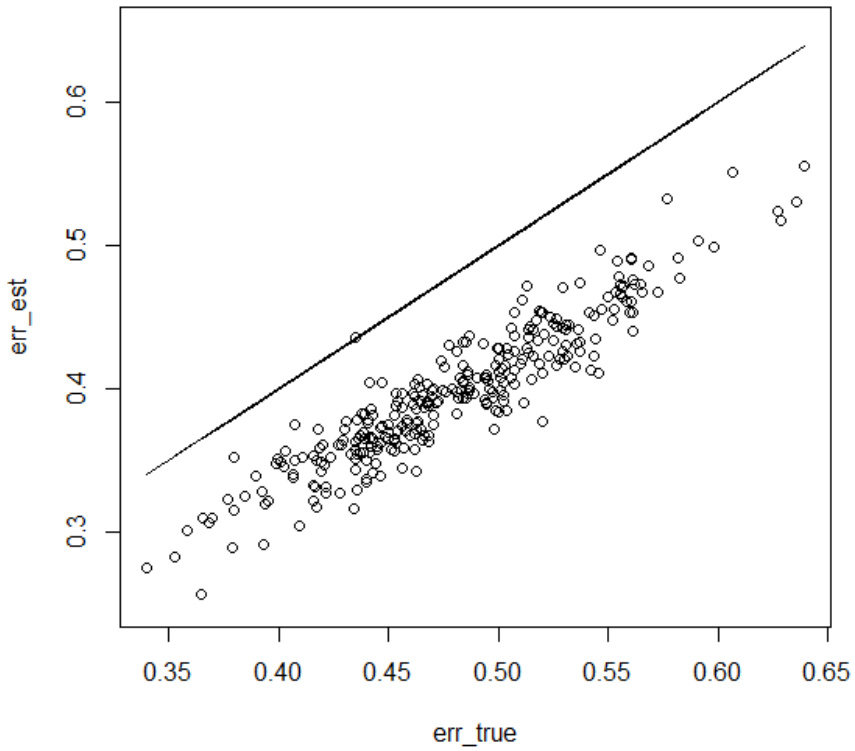
The Avg referred to would not be hard to evaluate by numerical integration, and the real value of this line of thought is for adjusting the distribution of estimated doses to correspond to the distribution of true doses. This was the most cumbersome part of the calculation of the current dose adjustments.

Making that simple has enabled the development of a simulation program to evaluate the performance of the proposed adjustment method.

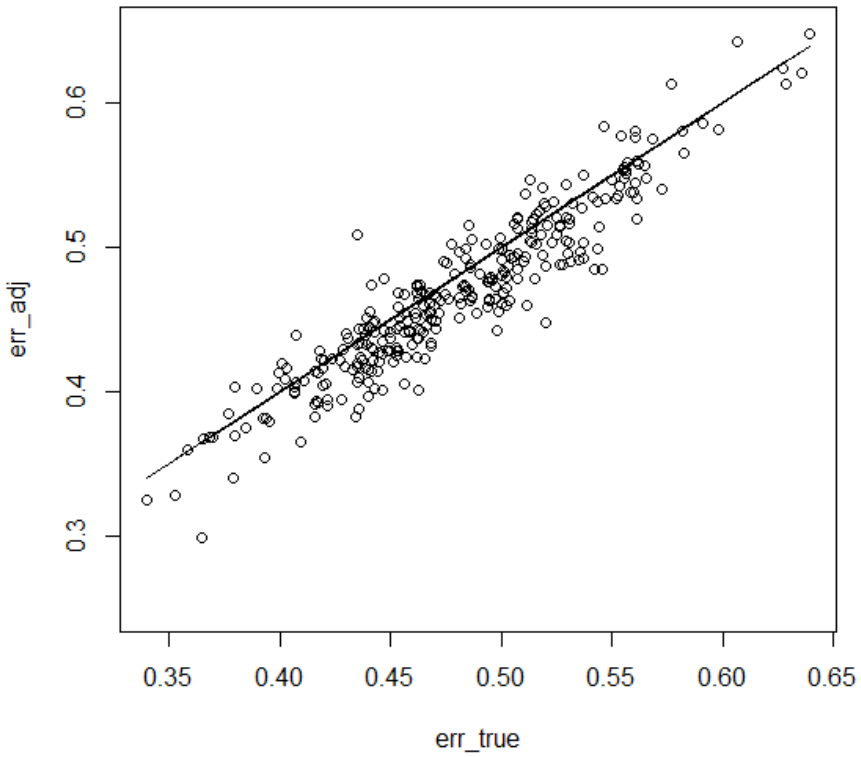
Idea is to take our dose estimates as though they were true doses, then add on errors under a model, generate cancer data under a true" model, then using all of "true", "estimated", "adjusted estimates" to carry out risk estimation for the known "true" model for the cancer deaths. This could never have been carried out with current dose adjustment methods.

Cancer risk estimates for 40% measurement and 20% averaging (Berkson) errors, where the assumed and true error levels are the same

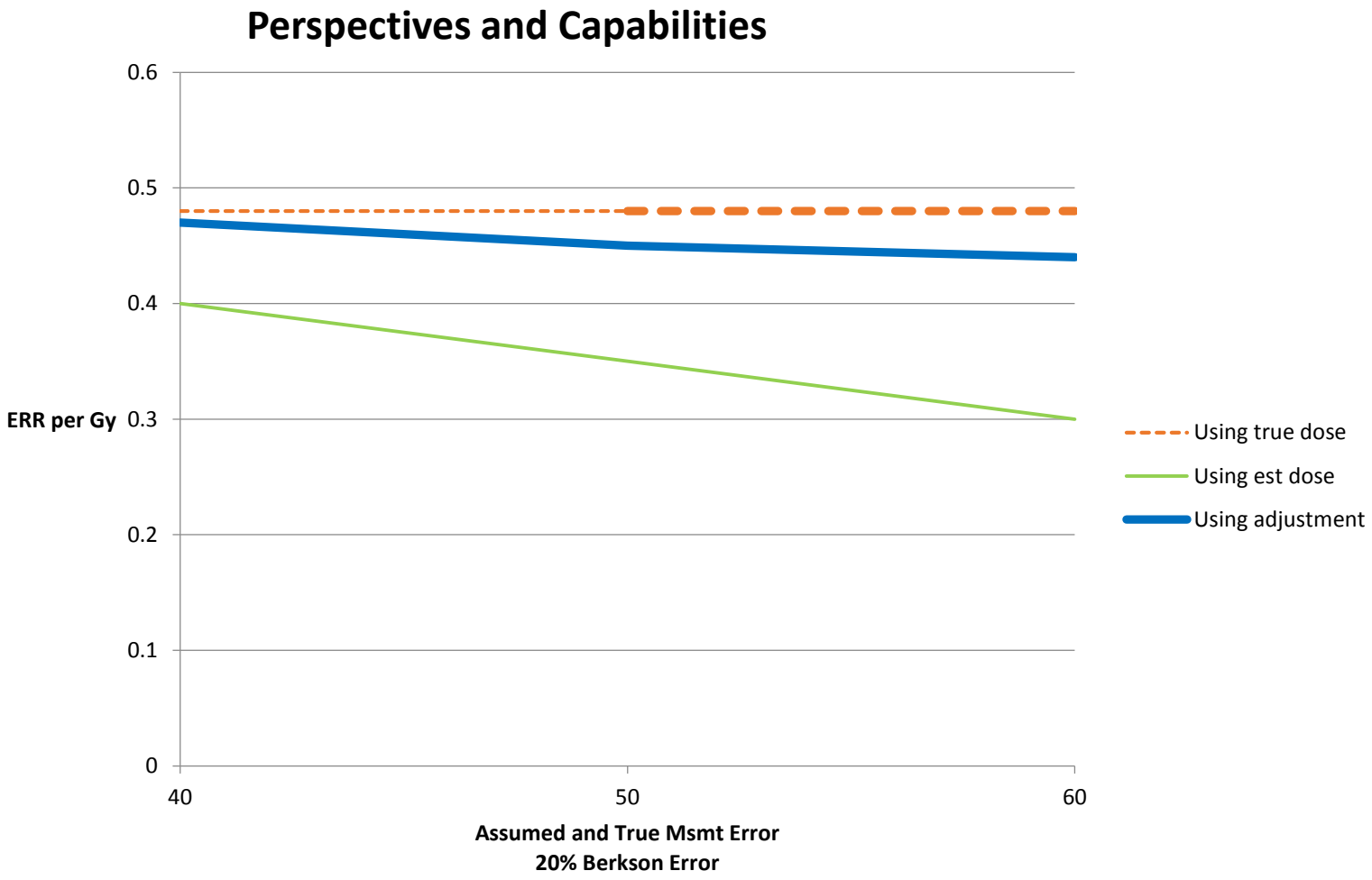
ERR using est dose



ERR using adj dose

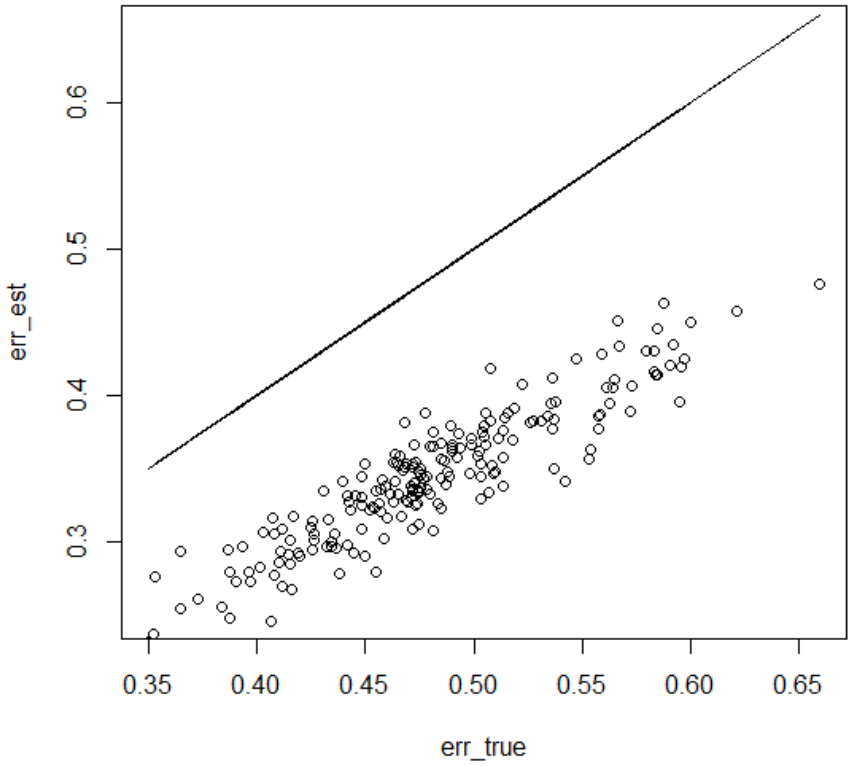


Same issue, presented differently and for a wider range of Msmt errors

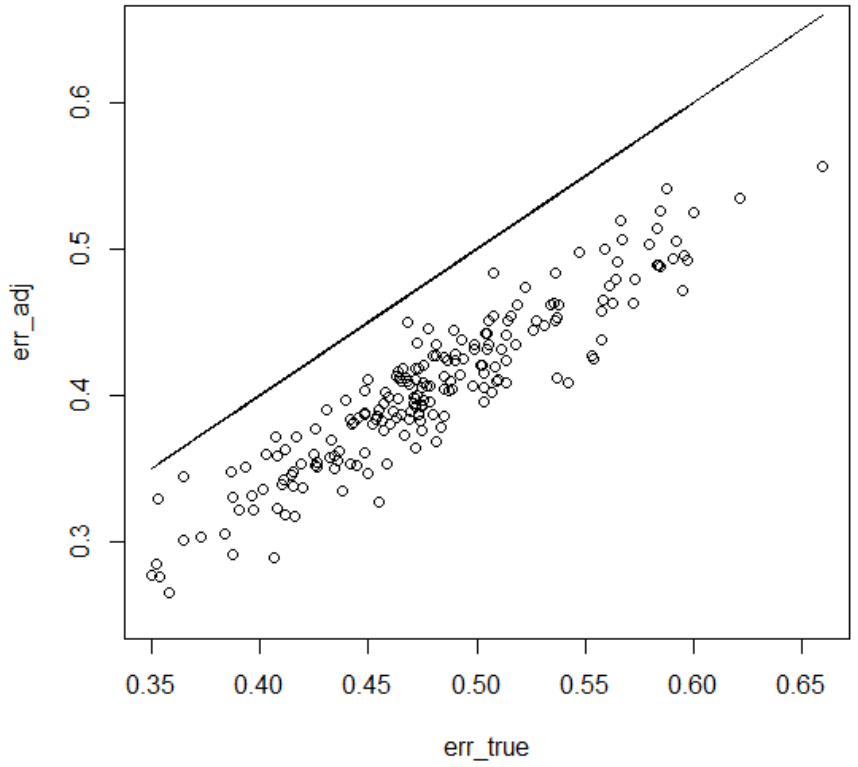


This is for assumed 40% msmt and 20% averaging, when the true values are 50% msmt and 00% averaging

ERR using est dose



ERR using *adj* dose



It is not only in computing this “Avg” where the local lognormal approximation is import to us.

In considerations above, we require an appximkation to the distribution of true doses, but we can only approximate directly the distribution of estimated doses. This matter was dealt with originally in a very cumbersome fashion. For the lognormal, or the local lognormal, these distributions differ by simple relation between the moments on a of log dose scale. These are simply carried over in the required functions *d1(z)* and *d2(z).*

Thus the proposed method provides in simple and more accurate calculations, readily automated, what was cumbersome in obtaining the current adjustments. One value of this is for doing simations as indicated next.

For current adjustments it was assumed that σ_M is 0.35, that is the SD of z is approximately 35% of the true dose. This is called the coeffiocient of variation CV.

In the proposed methods we can allow both for this measurement error, due to errors in assessing survivor location and shielding, and also a form of grouping (Berkson-type) error, where survivors in some shielding categories are all assigned a common shielding factor. Writing σ_G for the SD of log grouping error, the proposal is now to use values σ_M = 0.40 and σ_G = 0.20 , for a total SD of about 45% of the true dose.

Note that our main formula

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

Involves only σ_M , and the value of σ_G enters only for relating the distribution of true doses to that of estimated doses. (see 2008 ref paper)

We have reached some kind of landmark in important developments that began 30 years ago. The suggestions on how to implement this were published 8 years ago.

Part of my work this visit, and also in Oregon, is to indicate the change in cancer risk estimates that would result from moving from the current adjustments to the proposed ones. These changes are minor, and will not "upset any apple carts" regarding RERF cancer risk estimation. The value of moving to the modern methods is in being able to do explortions that could not be done using the current methods.

The other purpose of my visit and related work in Oregon, is to provide detailed documentation on how these modern methods can be implemented.

To be most clear, this involves computer code in R and Epiwin to carry out the implementation.