9 October 2009

NCI/NIAID Dosimetry Uncertainty Workshop: follow-up discussion

A few days after the Workshop, Owen Hoffman sent his impressions to a number of participants and asked them to respond to specific questions. The responses were very interesting and also led to another series of e-mail messages. The purpose of this document is to organize the e-mails that were exchanged in a logical manner. Some of the messages were edited to maintain the focus on issues related to the Workshop.

PART I: E-mails 1 through 9. Owen Hoffman's questions and participants' answers.

E-mail #1 **Date: 12 May 2009**
Author: Owen Hoffman **Topic: List of questions and comments**

Dear Colleagues,

I enjoyed meeting you last week, and wish we had much more time to engage in discussion. I felt that more "white space" was needed to clarify differences in terminology and concepts unique to each of our perspectives and disciplines.

I feel that many of you may have had the same after-workshop experience, and thus that it might be useful to follow up with an email summarizing some of my impressions and thoughts related to specific remarks that have stayed in my mind since last Friday's event. I will direct my comments and questions to specific speakers and/or members of the audience wherever possible.

1. **Roy Shore**: Important sources of uncertainty exist when exposure estimates depend on the answers provided in personal interviews. How are uncertainties in interview data considered in dose reconstruction models? I believe that errors entered into personal interviews will result in misclassification and bias in individual doses. When interviews are conducted decades after an exposure event, there can be considerable recall bias. Is this bias taken into account when uncertainty analysis is performed on dose reconstruction models? If so, how is it taken into account?

2. **Harry Cullings**: To what extent will I-131 in A-bomb fallout be an important factor to consider when analyzing the dose response for LSS thyroid disease? Could this not be a contributing reason why the most recent evaluation of autoimmune thyroiditis in the A-bomb survivors failed to show a statistically significant relationship between dose and disease?

3. **Don Pierce**: When you refer to a distribution of true dose, are you referring to a distribution of possibly-true dose to each individual in your cohort, or are you referring to a possibly-true distribution describing the variability of true doses among the individuals in your cohort? If it is the latter, then your probability distribution of possibly-true doses appears to be equivalent to what I refer to as alternative realizations of possibly-true distributions of cohort doses. Eduard Hofer refers to this as a set of "M" vectors of possibly true dose.

4. **Dale Preston**: I recognize the need for epidemiologists and biostatisticians to have an efficient and practical approach to evaluate the effect of uncertainty in dose or the dose response. However, when individual doses have very large uncertainties (the ratio of the upper and lower bound of a 95% uncertainty range in individual dose exceeds a factor of 10), and when the potential for varying degrees of shared bias in the mean dose estimate (per person and per cohort subgroup) is large, I see no other way to approach this problem other than to present the dose reconstruction results as a set of alternative realizations of possibly-true distributions of cohort doses. I believe this is what Eduard Hofer refers to as a set of "M" dose vectors.

   Each realization of a possibly-true distribution (vector) of cohort doses would be conditioned on a unique selection of a set of possibly-true parameter values that are shared among and within cohort subgroups. Each realization of a possibly-true distribution of cohort doses would also be conditioned on a set of possibly-true values describing residence history, dietary sources and dietary habits that would be unique to each individual in the cohort (i.e., uncertainties in interview data). Thus, each realization of a possibly-true cohort distribution would be conditioned on a unique set of possibly-true parameter values, some of which would be fully or partially shared and some of which would be unshared by individuals within and among cohort subgroups.

   I can envision the substitution of the individual mean dose as a surrogate for that individual's true dose as a useful approximation, (a) when systematic sources of uncertainties (shared sources of potential bias in model parameters and model structure) are small, (b) when the individual dose uncertainties themselves are small (i.e., 30% to 40%, or smaller), and (c) when the dose-response model is linear.

5. **Steve Simon**: I believe our two presentations complemented each other. I noticed that you use the term "trial" for what I refer to as an alternative realization of a possibly-true distribution (vector) of cohort doses.

   I would like to ask if you preserve all the model inputs and intermediate values that determine each alternative realization of a distribution of cohort doses. If not, then it becomes impossible to evaluate these alternative realizations for their "degree of plausibility," other than by simply weighting each by their goodness of fit to the vector of disease outcomes.

6. **Deukwoo Kwon**: Your analysis gave the impression that using the individual mean dose per person produced sufficiently adequate results when compared with the results produced by an explicit uncertainty analysis. To what extent is this result simply due to the small size of the uncertainty in your example? To what extent is this outcome the result of small amounts of uncertainties that are shared across cohort subgroups? How would your results change if your alternative realizations of distributions of cohort doses were to be increased from 100 to, say, 1000 or more?

   Do you expect the mean dose per person to produce adequate results when analyzing dose

reconstruction information produced from internal exposures (which should be much more uncertain than the doses from external exposures and for which varying degrees of potential bias would be more extensively shared among and within cohort subgroups)?

7. **Charles Land**: To what extent could uncertainty in dose explain the differences in the Kazakhstan dose response for internal and external exposure that was observed when the individual dose was simply represented by a mean dose and the effect of dose uncertainty was ignored? I believe the I-131 internal doses for the Kazakhstan cohort are quite uncertain, and this increased dose uncertainty may well explain the differences in the reported dose response between external and internal exposure.

8. **Vladimir Drozdovitch**: I enjoyed the fact that three independent groups have been used to verify the calculation of dose. Rigorous QA/QC of complex dose reconstruction models must be an essential prerequisite for epidemiological use of dose estimations and their uncertainty. How difficult was it for the three groups to obtain identical results (i.e., results that were free of programming and data errors)?

   For those model parameters described as frequency distributions representing stochastic inter-individual variability of exposure and dose, I would hope that the presence of uncertainty in the center and spread of these distributions is also taken into account. These are important sources of systematic and shared uncertainty that should not be overlooked.

9. **Elisabeth Cardis**: If only one alternative realization of a possibly true distribution (vector) of cohort doses shows a strong relationship with disease, and when this single realization is the outcome of a set of 10,000 realizations, I would indeed be highly suspect of the result. The chance is very high that this single good fit to the expression of disease cases is spurious. However, I would want to investigate the details of the dose model that produced this realization, to see if some unique combination of parameter values revealed information that could be used to judge the degree of plausibility of the result.

   In the example that I used in my presentation, there was a 90% chance that the wind blew to the north. However, 10% or less of the alternative realizations of possibly-true distributions of cohort doses might show a reasonable fit to the disease outcomes under the assumption that there is a 10% chance that the wind blew to the south at the time of the accident. For a situation using 10,000 alternative readings of possibly-true cohort distributions of dose, this would amount to 1,000 or fewer realizations showing a high goodness of fit, while 9,000 or more realizations would show a poor fit to the disease outcomes. I believe this number of "high fit" realizations would be sufficient to warrant re-investigation of wind direction.

10. **Alina Brenner**: It is true that numerous alternative realizations of possibly-true distributions (vectors) of cohort dose have the potential to present major difficulties to an epidemiologist who is accustomed to investigation of only one distribution of cohort dose composed of mean values or best estimates of individual dose. However, if completely different realizations of possibly-true distributions of cohort doses produced high goodness of fit to different disease endpoints, without any realization having a common condition or

parameter value, then this would be a clear indicator that the uncertainty in dose reconstruction (and in the epidemiological data) is too large to infer definitive conclusions about the underlying shape and slope of the dose response. I believe this will be an inherent problem with all epidemiological studies that rely on mathematical models to predict internal dose. This problem will be exacerbated when bioassay data, or data on body burdens or organ burdens, are not available to calibrate the environmental and dosimetric models that are used to predict internal dose.

11. **Eduard Hofer**: Could you explain your statement that "regression calibration" won't work in the cases where complex dose uncertainties are represented by a set of M vectors of possibly true doses? Is this because regression calibration may not be sensitive to large degrees of shared bias in model parameters and/or bias due to uncertainty in model structure?

12. **Tom Louis**: I enjoyed meeting you and listening to your perspective offered as a professional biostatistician. Could you explain why it is that the harmonic mean rather than the arithmetic mean is the preferred quantity to use when calculating a likelihood function?

    Several statisticians referred to the need to do preliminary sensitivity analysis on the uncertain doses. I am not totally certain of how I would do a preliminary "sensitivity analysis" when the sources of uncertainty in individual dose are numerous and complex (i.e., large, partially shared sources of bias combined with large sources of uncertainty that are unshared, along with other sources of uncertainty that represent large amounts of stochastic inter-individual variability).

13. **Andre Bouville**: I'm surprised that uncertainty in the milk transfer coefficient and in the weathering of I-131 from vegetation did not score high as important contributors to shared uncertainty among and within cohort subgroups. Could this be because these parameters are being treated as sources of unshared stochastic variability? Placing shared uncertainty on the geometric mean and geometric standard deviation of the distributions that describe stochastic inter-individual variability may take care of this problem, if it exists.

    To remove the effect of uncertainty in thyroid mass, it might be useful to analyze the dose response on the basis of organ burden (i.e., energy absorbed per organ) and not organ dose (i.e., energy absorbed per organ mass).

14. **Dan Strom**: You made excellent comments regarding difficulties in communication due to differences in the definition of terms and differences in the use of language to describe what it is we are talking about. It will take much more than a single workshop to resolve this issue. Hopefully, NCI will take a lead role, as they are the only organization I know of, other than perhaps RERF, in which dosimetrists, epidemiologists, and biostatisticians work together within the same organization under the same roof.

    Regarding your new term "dose swaggery" : you said that this term applies to dose reconstructions whose uncertainties exceed a factor of 20; (using the ratio of the upper and lower limits of a 95% uncertainty range, or dose uncertainties described by a lognormal

distribution with a GSD equal to or greater than 2.11). This condition describes nearly all of the environmental pathway dose reconstructions involving the use of environmental pathway models that I know of, including Utah, Hanford, Techa River, Chernobyl, Mayak, Kazakhstan and others.

When uncertainties in individual dose are this large or larger, and when they are the result of complex combinations of shared and unshared sources of potential bias in model parameters and model structure, I urge a healthy degree of skepticism when epidemiological analyses conclude that "uncertainty in dose had little to no effect on the slope of the dose response, nor on the width of the confidence intervals about the central estimate of the dose response."

I believe that in many instances, uncertainties in dose reconstruction may be sufficiently large and complex to preclude any definitive statement about the precise slope of the dose response. In most of these cases, the width of the confidence intervals about the point estimate of the dose response should be quite large.

Perhaps a conclusion from this workshop might be a recommendation for due caution when dose response analyses are based on dose reconstructions that contain uncertainties that are high and complex. Perhaps this is a reason why present quantitative risk estimates for radiation-induced cancer are based almost entirely on externally-exposed cohorts (with radon being a notable exception).

Once again, I enjoyed meeting everyone, and hope that my after-conference thoughts and questions may stimulate additional discussion.

Sincerely,

Owen

F. Owen Hoffman, Ph.D.
President and Director
SENES Oak Ridge, Inc.
102 Donner Drive
Oak Ridge, TN 37830
865-483-6111
senesor@senes.com
www.senes.com

E-mail #2 **Date: 12 May 2009**
Author: Roy Shore **Topic: List of questions and comments**

Owen,

I thought Andre had an excellent idea by organizing the workshop as he did to try to get input from various perspectives regarding a targeted set of studies. It didn't always work as fully as hoped, but at least we could talk to each other centered around a set of common problems.

As to your questions to me, some of them need to be directed to the dosimetrists who devise the dosimetric uncertainty models. As to misclassification & bias from interviews, for sure misrecall leads to misclassification error, and it may lead to bias (beyond dose-response attenuation) as well if there are innate tendencies to skew responses (e.g., people who eat a lot tend, in America at least, to underreport the amounts eaten), or if there are other reas ons for motivated bias in reporting (e.g., compensation issues, knowledge about their disease status). One of my slides was specifically oriented toward showing that in case-control studies, where there was knowledge of disease status, some evidence points toward there being reporting bias. In fact, it's something you might think about regarding the Utah fallout study, since the interviews were conducted after subjects knew their thyroid screening results. However, regarding your questions about taking interview uncertainties and/or bias into account in the dose reconstruction models, it's probably specific to an individual study & requires careful thought (hopefully by the epidemiologist, dosimetrist & statistician together) in planning and implementing the dose reconstruction.

Best,
Roy
shore@rerf.or.jp

E-mail #3 **Date: 13 May 2009**
Author: Harry Cullings **Topic: List of questions and comments**

Dear Owen,

I also very much enjoyed the workshop, and I wished it could have been longer. I hope it will be continued in some way, and perhaps reconvened much sooner than twelve years hence.

In regard to your point about I-131 in fallout in relation to thyroid outcomes in the atomic bomb survivors, I think it might be interesting to investigate whether geospatial methods could be informative in an epidemiological investigation of this. I would be worried about issues of statistical power to detect an effect, which would be difficult to pin down in the geospatial context but would be interesting to investigate.

Regarding the absence of a dose effect for autoimmune thyroiditis, I would again want to know whether there was enough power there to expect to see an effect, given the doses and numbers of person years involved and a reasonable alternative hypothesis in regard to the dose-response parameter(s). I must admit that offhand I don't know and would need to check on that.

I realize that one could attempt a dosimetric calculation for I-131 by assuming or estimating a fractionation ratio that would relate areal deposition of I-131 to the measured values that we have available, of e.g., gamma ray exposure rate, or in more limited cases, Cs-137 deposition. But where internal deposition of I-131 is concerned, it seems to me the environmental transport would be very difficult to reconstruct and would have very large uncertainty. In addition to the usual issues of environmental bioconcentration and variability of dietary intake, there also seems to be a major issue with the possibility of an intervening effect of extreme weathering very early after the bombings due to heavy rainfall. In ligh of all this, I wonder whether a major effort at dosimetry would be worthwhile at the present state of knowledge, in contrast to something as

simple as assuming a crude proportionality between the measurements just mentioned and the thyroid dose from radioiodines.

These are just my initial reactions and represent very limited knowledge on my part; I would certainly be open to continued dialog on this.

Best regards,
Harry
cullings@iea.att.ne.jp

E-mail #4 **Date: 12 May 2009**
Author: Don Pierce **Topic: List of questions and comments**

All,

I enjoyed the workshop, and was surprised when Andre called off the quest for consensus. I couldn't decide whether he thought matters were too controversial, or simply that a consensus would be hard to reach because of the varied issues.

As for Owen's question of me, i don't really know what he means by those two terms. When i wrote p(x), with x as true dose, I just meant the probability distribution describing the true doses in the cohort. That the true doses are unknown is not an issue there, since x is simply a variable to be evaluated at all possibilities. The notation p(x) is, however, not mathematically correct although common. a better way to write it is as p_X(.) , where the X is a subscript denoting the name of the random variable under consideration. that the distribution is not known exactly is an issue, and it has to be estimated, largely based on the observed distribution of estimated doses (with some correction to back out the errors).

I think that Louis and Spiegelman may have given the impression that integrating with respect to p(x | z, y) is more fundamental than i believe it is. To obtain the likelihood in terms of z (estimated dose), it is correct to integrate as

Int { L(beta ; x) f(x | z) dx }

without conditioning on the outcome y. In fact this is sort of the definition of the likelihood in terms of z. it is probably wrong to replace the likelihood there with the log likelihood or the profile likelihood.

To my knowledge the conditioning on y as well comes in for something like the following, and perhaps other ways. I have not verified what I give now, but this is my memory of things. The derivative of the log likelihood with respect to beta is called the score, and beta can be estimated by setting the score equal to zero, and solving for beta. The score can be computed by differentiating the logarithm of the above expression with respect to beta. But it also can be computed, sometimes more conveniently, as something like

Int{ d log L(beta ; x) / d beta f(x | z, y) dx }

where the conditioning on y comes in. If i am wrong in these things, Tom can correct me. the Stram & Kopecky paper has something related to this using likelihood ratios, which are essentially the same thing as the derivative of the log likelihood.

The first integral above can be approximated by sampling from f(x | z) and then averaging the values of L(beta ; x) from this. Again, it is probably wrong to average the log likelihood or profile likelihood in such a manner.

If there are only Berkson errors then f(x | z) is somewhat easier to deal with --- e.g. under usual conditions the mean of this distribution is z, contrary to what happens with classical errors.

I presume that sampling from the dosimetry system is far more closely related to dealing with Berkson errors than with classical. Stram & Kopecky say that one should deal with the classical errors first, making the "calibration", and then sample from the dosimetry system using these.

Best,

Don
pierce.don.a@gmail.com

E-mail #5 **Date: 13 May 2009**
Author: Dale Preston **Topic: List of questions and comments**

Owen

I too enjoyed last Friday's meeting and feel that many interesting points were raised. I understand the motivations and need for complex dosimetry systems, but I also feel that we are (or at least I am) a long way from understanding what these systems can and cannot address and from offering users a means to make practical use of a large number of "possibly-true-dose realizations" in model-development and other aspects radiation risk estimation.

Perhaps I am still missing something, but it seems to me that (as suggested by Stram and Kopecky and noted by Ethel in her nice talk last Friday) complex monte-carlo dosimetry systems deal with measurement (classical) error only to the extent to which the distributions for the parameters of the dosimetry system are corrected for measurement error (or bias) in the characterization of these distributions. For example, in your wind-direction example the problem is biased specification of the distribution of a crucial parameter of the dosimetry system, which I suspect would only become obvious with likelihood averaging or in a more fully Bayesian analysis (assuming there is enough data to dominate the prior assumptions about the dosimetry parameters). I have done a number of analyses and some (rather crude) simulations in which I have compared likelihood-averaging and the use mean doses derived from the realizations and for the most part have seen little difference between the risk estimates or their confidence intervals. I realize that there are examples (such as the thyroid neoplasm analyses described by Mallick et al 2007) where a fully Bayesian analysis that accounts for measurement (classical) and grouping (Berkson) error results in strikingly different estimates and inferences than a simple regression calibration analysis. However, I think that the comparison is not the same as comparing the Bayesian (MCMC or MCEM) methods to a simple analysis based on the mean of a sample of possibly true doses (which is what we are told these complex dosimetry systems are providing).

Perhaps I am being naive (certainly not for the first time), but I continue to think that a relatively straightforward approach in which one carries out basic analyses using mean doses from the M realizations of the dosimetry system followed by likelihood averaging or Bayesian analysis of a final model is a relatively practical method for dealing with dose uncertainty issues in complex problems. If the results disagree in any substantive way then one should consider what aspects of the dosimetry model are causing these differences.

As an aside, Owen commented that the use of mean doses would be problematic in non-linear

dose response models -- this could be addressed to some extent by using E(f(d)) as computed from the dose realizations in place of f(E(d)).

Dale
preston@hirosoft.net

E-mail #6 **Date: 26 May 2009**
Author: Charles Land **Topic: List of questions and comments**

Owen, I'm glad you raised your questions, and have benefited from reading the responses by the various participants. Just for completeness, the uncertain dosimetry estimates used in Deukwoo's analysis are still preliminary, and this is particularly true for the internal doses. In fact, the internal doses used were just placeholders generated by adding some uncertainty to the deterministic dose estimates used in our 2008 publication in Radiation Research. So I wouldn't draw any conclusions based on a comparison between risk coefficients for the external and internal doses.

Charles
landc@exchange.nih.gov

E-mail #7 **Date: 25 May 2009**
Author: Eduard Hofer **Topic: List of questions and comments**

Owen:

Please find my reply to your request:

"Could you explain your statement that ´regression calibration´ won't work in the cases where complex dose uncertainties are represented by a set of m vectors of possibly true doses?"

The outputs from a Monte Carlo uncertainty analysis of a dose reconstruction for n individuals are m possibly true vectors of n dose values each. It is important to note, that they can only be possibly true vectors if all sets of values used in the dose reconstruction model, that are subject to classical error, have been converted into m possibly true sets of values. These are sets of values with possibly true sample variance (the sample variance of the true values). Regression calibrated values are of no use here as one needs a sample of m possibly true sets of values. The true sample variance does not always have to be smaller than the sample variance of the measured values. If the sample correlation between true values and errors is negative, the set of measured values may have a sample variance that is smaller than that of the true values. Any correction blindly aiming at variance reduction will, in this case, even worsen the situation. One method for converting a set of measured values into m possibly true sets of values suitable for a Monte Carlo uncertainty analysis is described in (Hofer 2008). I wrote this paper specifically to show that this conversion can be done within the framework of an uncertainty analysis using Monte Carlo simulation. The uncertainty analysis is done properly only if all sets of values that are subject to classical error are converted into m possibly true sets of values that are then used in the Monte Carlo uncertainty analysis of the dose reconstruction. In this case there is no need for

any corrective actions, like regression calibration motivated by classical errors, at the level of the reconstructed dose values.

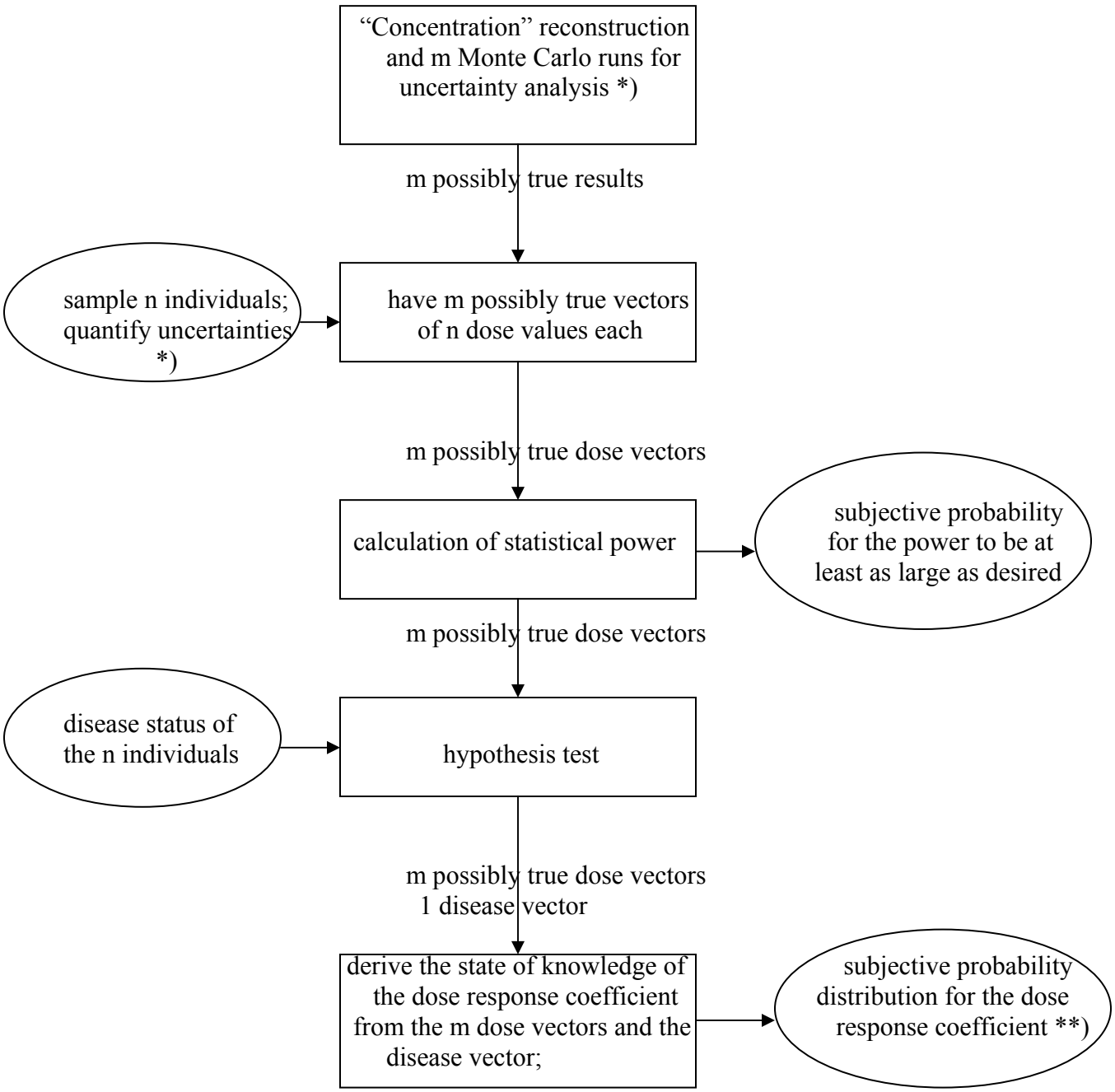
The Monte Carlo sample of m dose vectors are to express the state of knowledge of the n true dose values. Only one of these m vectors can be true. The state of knowledge follows logically from the state of knowledge quantified for shared and unshared uncertain parameter values, for input values subject to Berkson error, for input values subject to classical error, for uncertain functional relationships and, last but not least, by properly accounting for any state of knowledge dependences between these uncertain items of the dose reconstruction. It becomes immediately clear, that (unlike regression would assume) the m dose vectors, or the m dose values per individual, are not repetitions in the sense of a repeated measurement subject to classical error, i.e. varying about a true value, nor are they true values varying about a measured value as in the Berkson error case. Instead, they are m samples from a joint subjective probability distribution for the n true dose values. They are not a priori false like a measured value but are a priori true. Consequently, they have to be treated as such in a Bayesian approach to the estimation of the dose response coefficient, yielding a subjective probability distribution for this coefficient.

Using only the vector of the n averages of the m dose values per individual in the estimation of the dose response coefficient does not make any sense. It is, firstly, not justified to assume, that the vector of mean dose values is anywhere nearer to the vector of true dose values than any of the m vectors. In fact, the vector of mean dose values will most likely even be nonsensical as its components will be based on averaging over mutually exclusive modelling assumptions (i.e. wind directions, chemical decomposition or no decomposition, etc.). Taking the vector of median dose values does not make sense either as its components will most likely be obtained on the basis of mutually exclusive modelling assumptions (median values of individuals i and j obtained with different wind directions at the same time of the release, etc.).

What if the m sets of values used in the Monte Carlo simulation for a set of measured values subject to classical error are not obtained through conversion into sets of possibly true sample variance? In this case it is generally hopeless to try to rectify this shortcoming in retrospect at the level of the m vectors of n computed dose values. Elements of these improperly obtained sets of values will have entered the computation of different individuals' doses in differing numbers and in various functional forms.

Procedural Steps

The procedural steps outlined below are in my opinion the way to go:



*) sets of input values that are subject to classical error are to be converted into m possibly true sets of values;

**) this distribution quantitatively expresses the state of knowledge of the dose response coefficient as it logically follows from the influence of all quantified uncertainties of the dose

reconstruction as well as from the uncertainty in estimating the coefficient on the basis of the observed disease vector. From this distribution one reades Q% subjective probability intervals and P% upper and/or lower subjective probability limits for the dose response coefficient. The distribution may be the result of a Bayesian approach using the a priori density $f_0(\square)$ and the mixture of the m likelihoods $L(\mathbf{y} \mid \square;\mathbf{x_j})]$, where **y** stands for the disease vector, $\square$ for the response coefficient und $\mathbf{x_j}$ for the dose vector, j=1,...,m, such that

$$f(\square \mid \mathbf{y}) = [1/m \sum_{j=1}^{m} L(\mathbf{y} \mid \square;\mathbf{x_j})]\ f_0(\square)\ / \int [1/m \sum_{j=1}^{m} L(\mathbf{y} \mid \square;\mathbf{x_j})]\ f_0(\square)\ d\square$$

(where the factor 1/m cancels out) is the density function of the resulting subjective probability distribution for the dose response coefficient.

Best regards,

Eduard
ELHofer@t-online.de

Hofer E. How to account for uncertainty due to measurement errors in an uncertainty analysis using Monte Carlo simulation, Health Physics 95:277-290; 2008.

E-mail #8 **Date: 13 May 2009**
Author: Tom Louis **Topic: List of questions and comments**

Dear Owen,

I enjoyed the workshop and meeting you (and others!). I'll clarify my comment on the Harmonic mean issue. To strip the situation to its essentials, let \theta be the parameter of interest, Z the true dose, X the measured dose and Y the outcome. Then, the likelihood that needs to be maximized is:

L*(\theta; X, Y) = \int L(\theta; Z, Y) g(Z | X) dZ,

where g is the "dosimetry model." This is the "mean likelihood" (arithmetic mean, not harmonic). This likelihood can be maximized however you would like obtaining maxL* = L*(\theta*; X, Y), where \theta* is the MLE for L*. So, maxL* is the maximum value of what we really want.

Now, consider a multiple imputation approach. For this, one repeatedly generates "imputed doses" Z_i, i = 1, ..., M (I'll stay away from the issue of how to generate them!). For each we use L(\theta, Z_i, Y) to produce MLEs \theta_i, i = 1, ..., M and the related maximized likelihoods, maxL_i, i = i, ..., M.

In the standard approach the \theta_i are averaged to produce the overall point estimate. The variance of this estimate is estimated by the average of the imputation-specific variances (inverse curvatures of the imputation-specific likelihoods) PLUS the between imputation variation in the

\theta_i. However, sometimes you want an approximation to the maximized value of the actual likelihood, (i.e., maxL*). The arithmetic average of the maxL_i overestimates this value, which is quite intuitive because it is the arithmetic average of the individually maximized likelihoods as compared to the maximum of the arithmetic mean likelihood. It turns out that the harmonic mean of the individually maximized likelihoods provides a quite good approximation to maxL*.

I assume you are all happy that I didn't try to say this last Friday!

You will also find below my notes related to the workshop.

Regards,

Tom
Professor Thomas A. Louis, PhD
tlouis@jhsph.edu
http://www.biostat.jhsph.edu/~tlouis/

May 9, 2009 T. A. Louis

Dosimetry Notes

- Sociology: Need the dosimetrist, the statistician and the epidemiologist to be connected at the hip, from time 0

- Likelihood Averaging: Use the Harmonic mean, not the mean and not the geometric mean. Will give smaller likelihoods, lower information, higher variance. It is the best approximation to the full errors in variables likelihood.

- Errors in variables reduce information, can attenuate slopes, change shapes, even for the linear model.

- Dosimetry produces [D | X] (D is the true dose, X is the measured/estimated dose along with other information.

- Full Likelihood-based or Bayesian modeling, melds the dosimetry with the data likelihood.

- Multivariate Errors in variables can produce some apparent anomalies

- Using multiple imputation, don't summarize the output, use it to produce the full distribution of imputations. Don't need to assume overall normality even if the data likelihood is normal. Yes, the Rubin "within + between imputation" is the right way to get the overall variance, but why not use the full, mixture (of Normals) distribution to do the CIs?

- MI has an advantage over likelihood approaches in that you can use information for the imputations that can't be released in general. This was Rubin's motivation in the Census application.

- The "draws" from MI or from MCMC are weighted by a goodness of fit computation. See my mathematical attachments.

- Beware of Jensen's inequality. If you want to summarize a non-linear function of dose, e.g., T(D), then E(T(D)) \ne T(E(D)). If V(D)>0 and T is convex, then E(T(D)) > T(E(D)), if T is concave, the relation goes the other way.

- This underlies the presentation regarding power of a test (T is power). The computations were done in the convex part of the power curve. If they had been done for the concave region (power near 0.80) the relations would have gone the other way (D being variable would produce lower power). Power as a random variable, E(Power) or pr(Power > 0.80). You get lower power than the plug-in because you are brining in the true uncertainty. Of course, if you compare it to the wrong model, like the model with plugged-in median dose, you may do better. But, if you compare it to the "best" plug-in, you should do worse. Also, until the power is sufficiently high, Jensen's will work to make E(Power) > Power(Expectation). So, you need to be at the concave part of the power curve, that is around Z = 1 for the Gaussian distribution

- Big issue: Choice between using what is optimal for one endpoint or model or using a quite good, general-purpose approach that isn't globally optimal. Effectiveness depends on how informative is [D | X] relative to [D | X, Y, \theta] as computed from the likelihood. If Y doesn't matter much, or if just using it as a predictor, but not filtering through the analysis model, works very well. That is, $[D |X, Y]_{naive}$ \approx $[D | X, Y, \theta]_{correct}$
    - Strategy: Build a set of imputations on a very rich analytic model (lots of covariates, etc. and possibly include Y via the analysis model. Then, use it for lots of things, that model, sub-models, different model forms, ...

- Really good dosimetry, can make [D |X] sufficient for most tasks, without getting fancy with the fully Bayes model.

- Strictly speaking, need to sample from the hyper-prior or will get uncertainty wrong, and maybe shape.

- A, B and AB parameters, a fine typology, but might be easier to communicate and evaluate if discussed relative to a hierarchical model as represented by a directed graph.

- All of the EIV sophistication is important, but does not replace the need for valid and accurate dosimetry models, [D | X] or correct data likelihoods. Space age analyses will not rescue stone-age data!

- Be careful about treating all realizations with equal weight; CI issue from MCMC or just MC. Need the (smoothed) histogram and not just look at the range of the draws. The latter will give the impression that all have equal "density."
- You need a good analysis model even if you have the exact dose!

- Given a model, we do know what to do. Do we have to do it all exactly right, or can we approximate/shortcut in some places?

- Computer science enhancements: Importance sampling, rejection sampling, control variables, Metropolis-Hastings

- Need to incorporate and accommodate all sorts of correlations, within individual over time, between individuals over space and time, ..

.

- Reproducible research, documentation and software (from database to tables and figures)

- "Uncertainty" is not only spread, but also shape (e.g, bimodal). Really, the goal is to get the full predictive distribution and the "do something with it." The tails usually matter.

E-mail #9 **Date: 10 September 2009**
Author: André Bouville **Topic: List of questions and comments**

Owen,

1 – The reason why uncertainties in the milk transfer coefficient and in the weathering of I-131 from vegetation did not score high as important contributors to shared uncertainty (in the thyroid dose estimates) among and within cohort subgroups is because all individual thyroid doses in our Chernobyl cohort studies are based on direct thyroid measurements, which provide direct estimates of the I-131 activity in the thyroid at the time of measurement. Therefore, only the shape of the curve representing the temporal variation of the I-131 activity is needed. Because the shape of that curve does not depend on the absolute value of the milk transfer coefficient and the effective half-time of weathering of I-131 from vegetation is relatively well known, those two parameters do not influence much the uncertainty in the thyroid dose estimate.
2 – I do not see the necessity of removing the effect of uncertainty in the thyroid mass. Because our goal is to estimate the uncertainty in the individual thyroid dose estimate, the thyroid mass has to be considered at one time or another.

Best regards,

André
BouvillA@mail.nih.gov

PART II: E-mails 10 through 18. Follow-up discussion and other comments.

E-mail #10 **Date: 14 May 2009**
Author: Ethel Gilbert **Topic: Comments**

Owen, Andre, and others,

I too found the workshop very interesting, and hope that it will lead to a better understanding of the issues and eventually lead to approaches for at least get us closer to being able to correct for the distorting effects of dosimetry uncertainties. I think a summary describing some of the issues that were raised would be helpful.

As a result of the thinking I did to prepare for my talk, I have become reasonably convinced that although multiple realizations (such are done for thyroid dose reconstruction) are very helpful in addressing shared errors, they cannot by themselves address the attenuation in the dose-response brought about by unshared classical error. As both Don and Dale have noted (also Stram and Kopecky and Stayner et al.) simulations are based on the assumption of a Berkson framework, which means that any needed calibration has already been done. This is important since to my knowledge, none of the multiple realizations discussed at the meeting did any regression calibration to account for classical errors. However, I certainly don't rule out the possibility that I'm the one who is confused. In this case, I hope that someone can help me to understand how the simulations can address the attenuation resulting from classical errors.

It might help to consider a very simple example. Suppose that one has a device for measuring dose, and that the distribution of the measured dose conditional on the true dose is normal with a mean that is the true dose and a variance of V. Also assume that the errors are independent for different subjects. This is an unshared classical error so that without adjustment, the estimated slope of a linear dose-response will be biased toward the null. Now suppose that Monte Carlo simulations are carried out starting with the estimated doses, using the known V, and assuming independence. Although appropriate use of these simulations could increase the variance of the estimated regression coefficient, I can't see how they could be used to correct the attenuation (in fact they might lead to more attenuation because of the additional variability). As noted above, the simulation approach implicitly assumes that the errors are Berkson, and that regression calibration needs to be done prior to the simulations. Performing the needed regression calibration for complex dosimetry systems would be quite challenging and would certainly require input from both statisticians and dosimetrists. Although we are unlikely to be able to do this perfectly, perhaps we can develop approaches that could at least improve the status quo.

I'd also like to comment on the meaning of the expected value of the true dose conditional on the estimated dose (denoted by E(true|est)), which is what one needs to substitute for the estimated doses when performing regression calibration. The concept of E(true|est) makes sense only within a specified population. The concept is not useful for estimating dose for only one individual (or, in this cases E(true|est) would just be the estimated dose). For the regression situation, one can imagine a response y plotted against the true doses. With classical error, these doses are stretched out horizontally, thus flattening the dose-response. Regression calibration brings the doses back toward the center horizontally, thus correcting the flattening and leading to the correct slope of the dose-response. It may also help to think of drawing information from the

group in that higher doses are more likely to be overestimated than underestimated, whereas the reverse is true for lower doses. Regression calibration "shrinks" the doses toward the mean. I would recommend that anyone who is confused about this issue read the first few pages of Pierce, Stram, and Vaeth, Radiation Research, 1990.

Ethel
gilberte@mail.nih.gov

E-mail #11 **Date: 14 May 2009**
Author: Sarah Darby **Topic: Response to Ethel**

Hi Ethel

I was so sorry to have to miss the Workshop last week, as I was unwell. However I have been following the correspondence with interest. My paper with Tom Fearn and Dave Hill addressed the attenuation resulting from classical errors using simulation (reprint attached). Does this help you at all?

Best wishes

Sarah
sarah.darby@ctsu.ox.ac.uk

E-mail #12 **Date: 14 May 2009**
Author: Ethel Gilbert **Topic: Response to Sarah**

Hi Sarah,
You were missed at last week's meetings – I'm sure you could have made a strong contribution. I hope you are well now.

Thanks for sending the paper. I haven't yet read it carefully but it certainly seems to require information on the distribution of the true doses conditional on the estimated doses. The simulations I was commenting on in my e-mail start with the observed dose (or more generally several variables that are used in calculating dose) perturb these and conduct many realizations of doses for the population of interest (taking account of the correlation structure). (See Stram and Kopecky, or Stayner et al.) If my understanding is correct, there is no explicit consideration of the distribution of the true dose conditional on the estimated dose. What I meant to say in my e-mail is that I don't think this particular type of simulation can address classical error although simulations may play a role in other approaches that might be used. My understanding is still limited so it's good to have this dialogue.

Best wishes,
Ethel
gilberte@exchange.nih.gov

E-mail #13 **Date: 20 May 2009**
Author: Owen Hoffman **Topic: Additional thoughts**

Dear Colleagues,

I really appreciate the responses that I received last week in my attempt to encourage an after-workshop electronic discussion to follow up on numerous loose ends resulting from the presentations and somewhat limited discussion that occurred during the workshop itself. I have a few additional thoughts to share as a result of correspondence received from Ethel Gilbert, Dale Preston, and Roy Shore.

1. **Two-dimensional Monte Carlo presentations of uncertain dose reconstruction results vs. a general regression calibration on expected values of individual dose.**

The advantage of 2D-MC is to separate out the potential for systematic bias in fixed quantities (that are totally or partially shared among and within cohort subgroups) from processes that lead to stochastic (unexplained random) inter-individual variability of true dose. Ethel expressed the concern that each alternative realization of a vector of true doses likely contains a residual element of classical measurement error. The presence of classical error within a vector of true doses would pull the resulting dose response towards the null.

I wonder if there would be merit to an analysis in which alternative regression calibrations would be made, each conditioned on a single vector of possibly unbiased expected values of dose per person. This would involve expanding the calculational effort required of a 2-D Monte Carlo uncertainty analysis, so that alternative vectors of possibly unbiased mean dose per person would be obtained. Each vector of possibly unbiased mean dose would be conditioned on the selection of a unique set of possibly true fixed (but uncertain) parameter values. These are quantities for which only one value is possibly true for the exposure situation being evaluated. These possibly true fixed parameter values include quantities that are shared among and within cohort subgroups and those representing uncertainty in the residence histories and dietary habits unique to each individual in the cohort.

Thus, in the case in which there were 1,000 alternative vectors of true dose, the 2D-MC would be expanded to produce 1,000 alternative vectors of possibly unbiased individual mean doses. Regression calibration would then be performed on each vector of possibly unbiased individual mean doses.

In the example I used, in which wind direction was uncertain but present-day meteorology suggests a 90% probability that the wind blew towards the North and a 10% probability that the wind blew to the South, 900 of the alternative vectors of possibly unbiased mean doses would be generated for the condition in which the wind blew towards the North, and 100 of the alternative vectors of possibly unbiased mean doses would be produced for the condition in which the wind blew towards the South. Regression calibration would be performed on each of these vectors to eliminate residual degrees of classical measurement error inherent when simulating processes representing stochastic (unexplained) inter-individual variability of exposure and dose.

I know that Eduard Hofer has written a publication about how to address and eliminate classical measurement errors when a 2D-MC uncertainty analysis of a cohort dose reconstruction is conducted. I hope he will weigh in further on this point.

2. **When can general regression calibration be useful within the context of 2D-MC dose reconstruction?**

I believe that the only time in which general regression calibration (conditioned on one value of individual mean dose per subject obtained across all M realizations of vectors of true dose) would be useful is when the potential for systematic bias in model parameters (i.e., shared among and within cohort subgroups) is small and/or when the number of cohort subgroups for which shared sources of uncertainty are independent from group to group is large. In this latter case, shared sources of uncertainty would occur only within but not among cohort subgroups.

When there is a high degree of uncertainty associated with model parameters that could be biased and when this uncertainty is a source of systematic error that is shared within and among cohort subgroups, then it is difficult for me to see how a general regression calibration on a single mean dose per subject would be useful. Substitution of the individual mean dose by the vector of true doses that represents the average vector over all M vectors of true dose (as described by Dale) would offer no advantage in this case.

3. **The potential for uncertainty in dose due to bias in interview data**

Roy Shore mentioned in his previous response that the Univ. of Utah study of children exposed to fallout from the Nevada Test Site was biased because individuals who were interviewed were already aware of their disease status. I have received the following response from Ms. Mary Bishop-Stone of the Univ. of Utah. Here is what she says about this issue:

> "The cohort was examined in 1985 and 1986 and the parents who provided the amount and source of milk and vegetables were interviewed in 1987. It is, therefore, possible that the parents knew their child's disease status before they were interviewed. Naturally we had planned to do the interviewing before the examinations. This just didn't work out because of delays in getting the computer questionnaire ready. These things happen in research. However, we did two things to find out if this was a problem. We conducted a small telephone interview to determine how much people in Washington County knew about their exposure. It turns out that people were not generally aware that their exposure came through milk. We thought that the exposure route would be obvious, but it turned out not to have been obvious at the time to the general citizenry of Washington County. Secondly, we looked at the amount of milk reported by parents by county, stratified by disease status, and found that, on average, the parents in Washington County did not report more milk drinking than parents in the other two counties.
>
> It would have been virtually impossible for parents to have systematically exaggerated their children's milk drinking by dose status because neither they nor their children knew their calculated dose until the funding ended in August 2005, nearly 20 years after they had already completed the questionnaire. Some of the parents thought they knew that

> their children were exposed because of where they lived, but if they were unaware of the route of exposure then they could only have exaggerated based on residence information (which would have resulted in random misclassification because they did not really know which locations were associated with high versus low exposure)."

In addition, I have received the following from Dr. Lynn Lyon, also of the Univ. of UT, on the same subject:

> "The key determinant of exposure was, 'What was your milk source in the spring of 1953?' for those under age five at the time. This information came from subject's mother, not the subject. The amount of milk consumed was not the key determinant of increased risk of thyroid disease. It was the source of the milk, specifically milk from backyard cows and goats, that put people into the high risk group."

Best regards,

Owen
senesor@senes.com

E-mail #14 **Date: 15 May 2009**
Author: Owen Hoffman **Topic: What should we do next?**

Dear Andre,

I was wondering whether the electronic discussion that we are having subsequent to last Friday's workshop is information that might be worth sharing with all workshop participants. At present, I have only corresponded with the other presenters and Mary Schubauer-Berigan. I've had off-line discussions with Dale and Ethel, which I hope to be able to share with everyone once I get feedback confirming the merit of my ideas for addressing inadvertent residual levels of classical error in a 2-D Monte Carlo approach. Hopefully, Eduard Hofer will weigh in on some of this discussion. Much of this post-workshop discussion relates directly to the essence of his two short presentations.

In general, I think we are coming very close to reaching a consensus. Much of the remaining concern about 2-D Monte Carlo has to do with the need to correct for the presence of residual classical uncertainty in our Type A simulations of inter-individual variability. But there is a need to preserve alternative realizations of Type B uncertainty when we know that there is the potential for model parameters and model structure to be systematically biased among and within cohort subgroups. When uncertainty is dominated by inter-individual variability of true dose and when systematic sources of bias can be considered negligible, regression calibration on the arithmetic mean dose per individual (averaged over all realizations of all vectors of the cohort dose) should provide reasonable results. The problem is that when models are used as surrogates for direct measurements, it will be difficult to justify an absence of sources of systematic bias. Thus, there is a need to address these potential sources of bias explicitly within a 2-D Monte Carlo approach.

A remaining issue, then, is how best to weight each dose response that has been conditioned on each alternative vector of poss ibly true individual dose? I've told Dale and Ethel that if the presence of residual amounts of classical error is that big of a problem (in each simulation of a cohort dose vector), then a solution might be to simply estimate the possibly unbiased mean dose per person conditioned on each dose vector.

That's it for now. Most of what I really wanted to say was in the first paragraph. Thanks for inviting me to be involved with this discussion.

Owen
sensor@senes.com

E-mail #15 **Date: 14 May 2009**
Author: André Bouville **Topic: Preparation of a "consensus" paper?**

Dear colleagues,

it seems to me that I was too hasty when I proposed not to prepare a paper on what we learned during the uncertainty workshop. We certainly are not ready to prepare a "consensus" paper, but one describing the issues and the possible ways to address those issues may be timely. Owen, Steve, and I will try to find the time to prepare a dosimetry-oriented draft that the statisticians could build on. Please let me know if you think that it is a bad idea before we proceed.

Best regards,

Andre

E-mail #16 **Date: 14 May 2009**
Author: Colin Muirhead **Topic: Response to André**

Dear Andre,

This is a good idea. Like the others who responded, I enjoyed the workshop. It would be helpful to build upon it, by summarising the key issues for a wider audience.

Best regards,
Colin
Colin.Muirhead@hpa.org.uk

E-mail #17 **Date: 14 May 2009**
Author: Don Pierce **Topic: Disagreements?**

All,
I suspect that there might be a fair amount of disagreement on the role of using simulated dose estimates. To lay this out even without reaching a consensus would be useful, but then it might also be rather hard work (and maybe even a bit unpleasant). There is also the matter that the uncertainty issues cover a lot of territory that might be difficult to organize.

Not sure what I would recommend about attempting this, but I am willing to talk about it.
Best,
Don
pierce.don.a@gmail.com

E-mail #18 **Date: 02 June 2009**
Author: Peter Jacob **Topic: UNSCEAR plans**

Dear Andre,

I am glad that the workshop has been so successful and stimulated quite a bit of discussion and e-mail exchange after the workshop. It will be very helpful to have the presentations on the web and to publish a summary publication.

Presumably you are aware of the UNSCEAR plan to write a report on 'Uncertainties in radiation risk estimation'. This activity is in a very early stage. In the moment, just some elements for a project development plan are existing, which I send you confidentially in the attachment. I intend to elaborate the plan a bit in the next ten days. Any comments from the workshop participants would be very welcome. It is intended to establish a group of experts, who will commonly write the draft report.
Proposals for possible members of that expert's group are also welcome.
A sub-group of UNSCEAR will meet in July and discuss the issue.

Best regards,
Peter
Jacob@helmholtz-muenchen.de