DECAY OF ENVIRONMENTAL ^{137}Cs

Dear Editors:

A recent paper by Palms et al. (2007) showed environmental levels of ^{137}Cs in periphyton over time. Covering 25 years of monitoring data, the slow decay of this nuclide was evident. Health physicists have been detecting this nuclide since fission products were injected into the atmosphere by weapons testing. Many, including myself, have seen the levels decline. The robust data set in the Palms article allowed for a quantitative assessment of this decline.

I plotted the 25 years of ^{137}Cs data and fitted an exponential curve to them. This is shown in Fig. 1. The decay constant obtained is 0.0753 y^{-1}. This decay constant represents an effective half-life of 9.2 y, much less than the 30-y radioactive half-life of this nuclide. Although there is a confidence interval around this value, I did not attempt to estimate it.

A paper by Robison et al. (2003) specifically looked at the effective half-life of this nuclide in Pacific island tree leaves. Using concentration measurements spanning 36 y, they obtained a value of 8.5 y (95% confidence interval: 8.0 to 9.8 y). Considering that this ecosystem, a coral atoll, is vastly different than the temperate ecosystem studied by Palms, the values agree remarkably well. I would be interested to see if other data sets confirm this effective half-life.

Robison et al. attribute the loss of ^{137}Cs (non-radiological) to transport from soil to groundwater, removing it from the root zone of the trees. Perhaps a similar removal mechanism is at work in the Palms data.

JOEL I. CEHN

1036 *Hubert Road*
Oakland, CA 94610
cehn@aol.com

REFERENCES

Palms J, Patrick R, Kreeger D, Harris C. 25-y study of radionuclide monitoring with terrestrial and aquatic biomonitors. Health Phys 92:219–225; 2007.

Robison WL, Conrado CL, Bogen KT, Stoker AC. The effective and environmental half-life of ^{137}Cs at Coral Islands at the former US nuclear test site. J Environmental Radioact 69:207–223; 2003.

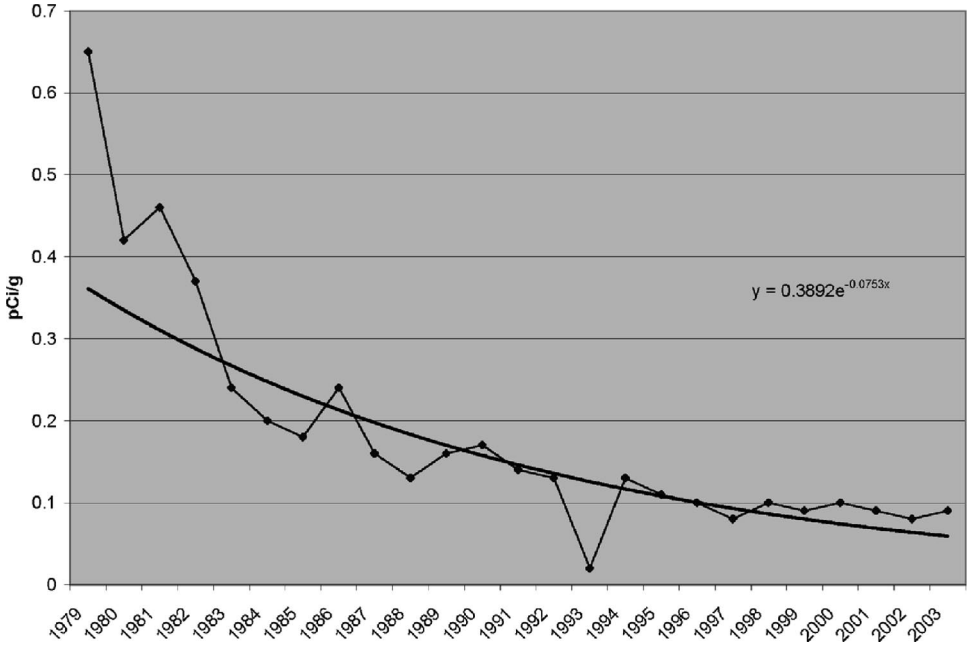

Fig. 1. ^{137}Cs in periphyton; data from Palms et al. (2007).

0017-9078/07/0

HYPOTHESIS TESTING, STATISTICAL POWER, AND CONFIDENCE LIMITS IN THE PRESENCE OF EPISTEMIC UNCERTAINTY

Dear Editors:

We are writing in regard to an article published in the March 2007 issue of *Health Physics* by Eduard Hofer (Hofer 2007) on the interesting and important subject of dealing with subjective uncertainty in radiation dosimetry when dose estimates are applied to epidemiological studies, especially when a sequence of alternative dose estimates rather than a single "best estimate" of dose is provided for the epidemiological application. We have described several approaches to this problem (Stram and Kopecky 2003; Kopecky et al. 2004), and while these may not provide the last word on this problem we have an important concern regarding the proposed analysis of Hofer, namely the type I error (or "false positive" error) properties of the proposed analysis.

Our approach (Stram and Kopecky 2003) and that of Hofer are similar in that both start with an assumption (often requiring a considerable leap of faith) that the dosimetry system used to estimate dose for each individual in the study can be regarded as providing estimates from a distribution of true dose conditional upon what is known about the determinants of the actual exposure. Specifically, both Hofer and we assume that the dosimetry system generates m independent sequences (or "replications") of dose estimates $\{x_{i,j}\}$ ($i = 1, \ldots, n$, is the index for individuals in the study and $j = 1, \ldots, m$ for the sequence number) for the n subjects from the conditional distribution of true dose given all that is known about the parameters, source terms, individual data (excluding outcome data), etc., determining the true exposures. In our 2003 paper, we described some of the operating characteristics of treating the sequences in a manner analogous to what is done in the so called "Berkson error" problem (Thomas et al. 1993). Specifically, we described some statistical implications of using the mean, $\{z_i\}$, of true dose given "all that is known" about true dose as the dose variable in a linear regression analysis relating disease to exposure. (We may estimate $\{z_i\}$ by averaging the m sequences, $\{x_{i,j}\}$, over j, assuming that m is large enough so that the estimation of this mean is very accurate.) Thus, for example, we reject the null hypothesis of "no exposure effect" in this analysis only if standard statistical tests (ignoring dosimetry error) concluded that there was an association between the mean doses, z_i, and the outcome of interest, Y_i, with the appropriate degree of confidence.

Hofer (2007) suggests a different test of the null hypothesis (this is most clear from the simulation experiment performed to compute power given in the Appendix of that paper), namely to use each replication $\{x_{i,j}\}$, $j = 1, \ldots, m$ in turn in m separate regression analyses (regressing Y_i on each sequence separately) so that a total of m tests are performed at a specific type I error rate (denoted as α). Then Hofer suggests (point 2 on page 233–234) rejecting the null hypothesis if more than 100α percent of these m separate regressions give a significant p-value ($p < \alpha$).

The problem with the proposed procedure is that it doesn't properly control the false positive rate, i.e., the type I error, α, of the test. That is, the new procedure will reject a true null hypothesis more often than 100α percent of the time. To see this, consider the special case when both the null hypothesis is true, i.e., that disease, Y_i, and true dose are independent of each other, and when $x_{i,j}$ and $x_{i',j}$ are also independent over the replications j. (The second assumption would hold when there is no information at all about true individual dose in the output from the dosimetry system.) In this case, a count of the number of times that the p-value is less than α (R_m, say) will be distributed as a binomial random variable with rate parameter α and m as the number of trials (since each sequence of x is independent and related only by chance to disease). As m increases to infinity the false positive rate of the procedure will therefore approach ½. (To see this, note that a false positive result from Hofer's proposed test corresponds to $R_m > \alpha \times m$, and that for sufficiently large m, R_m is approximately normally distributed with mean $\alpha \times m$.) For smaller m the actual false positive rate will still be considerably greater than the desired rate α. (For example, with $\alpha = 0.05$ and $m = 100$ the expected false positive rate is 38.4 percent.)

Dropping the assumption that $x_{i,j}$ and $x_{i',j}$ are independent over the replications j decreases the type I error of the proposed procedure, but it will remain inflated so long as these variables are not perfectly correlated (i.e., when there is no dosimetry error). Indeed, in the simulation experiment presented in the Appendix of Hofer (2007) we see that the false positive rate using $\alpha = 0.05$, while not 38.4%, was 13%, far greater than the 5% required for a test to be valid, while the error using the Kopecky (2004) approach was 3% (not statistically different than the desired 5% given the number of simulations performed). The conclusion made at the end of the Appendix, that the proposed procedure is more powerful than that of Kopecky et al., cannot be trusted because it ignores the overwhelming evidence that the proposed procedure is anti-conservative under the null.

In Stram and Kopecky (2003) we described several other candidate approaches to dealing with dose uncertainty, specifically when (as with the Hanford dosimetry) there are errors that are "shared" over many subjects. The method that is nearest in spirit to Hofer's proposal is Monte Carlo maximum likelihood (MCML). In this method the likelihood function itself is averaged over a large number of replications, *m,* and then maximized (with respect to its parameters) to find estimates and confidence intervals for the dose-response parameters of interest. This procedure, while computationally intensive, does show promise in dealing with errors in dosimetry systems that include both shared and unshared components. Further statistical work on this interesting and challenging problem is encouraged.

Acknowledgments—This work has been supported in part by grants R01 OH11869 (NIOSH), 5P30 ES07048 (NIEHS), and DE-FC01–07HS07010 (DOE).

DANIEL O. STRAM
DUNCAN C. THOMAS
Division of Biostatistics
Keck School of Medicine
University of Southern California
Los Angeles, CA 90033

KENNETH J. KOPECKY
Division of Public Health Sciences
Fred Hutchinson Cancer Research Center
Seattle, WA 98109

References

Hofer E. Hypothesis testing, statistical power, and confidence limits in the presence of epistemic uncertainty. Health Phys 92:226–235; 2007.

Kopecky KJ, Davis S, Hamilton TE, Saporito MS, Onstad LE. Estimation of thyroid radiation doses for the Hanford thyroid disease study: results and implications for statistical power of the epidemiological analyses. Health Phys 87:15–32; 2004.

Stram DO, Kopecky KJ. Power and uncertainty analysis of epidemiological studies of radiation-related disease risk where dose estimates are based upon a complex dosimetry system: some observations. Radiation Research 160:408–417; 2003.

Thomas D, Stram D, Dwyer J. Exposure measurement error: influence on exposure-disease relationships and methods of correction. Annals of Public Health 14:69–93; 1993.

RESPONSE TO STRAM ET AL.

Dear Editors:

The concerns voiced by Stram, Kopecky, and Thomas in their letter to the Editors of *Health Physics* do not invalidate the procedure suggested in my paper.

When faced with epistemic uncertainties of dose reconstruction:

- averaging over a sample of dose vectors,* where every single one could be the true vector, does not make sense;
- one can only obtain a subjective probability distribution for the statistical power of hypothesis tests; and
- the test result cannot be as precise as "reject the null hypothesis (H_0)" or "do not reject H_0" but can only be in the form of a subjective probability for rejection.

At the end of the day, of course, one needs to either reject or not reject H_0. An estimate of the subjective probability for rejection is obtained as the fraction of the m alternative dose vectors, produced by the uncertainty analysis of the dose reconstruction, that lead to rejection of H_0. If this fraction is larger than a prescribed value δ then it makes sense to conclude that H_0 should be rejected at the present state of knowledge of the dose values. Which value to use for δ may be a matter of debate. I chose $\delta = 0.05$, which happens to be the same value as for the significance level α of the hypothesis test with each single dose vector.

Stram, Kopecky, and Thomas criticize that in my example H_0 was rejected for 13 out of $m = 100$ disease vectors generated under the null hypothesis although the significance level (or type I error probability) of the test with each single dose vector was $\alpha = 0.05$. Given H_0 is true, then on average 5% of the disease vectors generated under H_0 will lead to rejection of H_0 for a given dose vector. If the m dose vectors are not very different, then H_0 will be rejected for mostly the same disease vectors irrespective of which dose vector is used in the test. But in my example (and in the Hanford Thyroid Disease Study), the m dose vectors are quite different from each other (i.e., dose uncertainty is large, see Fig. A1 of my paper). Consequently, the disease vectors leading to rejection of H_0 for a given dose vector will not be the

* A dose vector is an array of dose values assigning one dose value each to the individuals in the cohort.

same for each of the m dose vectors. Therefore it is not surprising that there are more than 5% of the disease vectors leading to rejection of H_0 for a fraction larger than $\delta = 0.05$ of the dose vectors.

It is quite easy to see that one can choose a value of δ larger than α such that the probability of rejection of a true null hypothesis does not exceed α. This will lower the subjective probability for the power of the test to be at least as high as required. Since it is often not possible to increase this subjective probability sufficiently by increasing the number of individuals in the cohort, there is no way out of this dilemma other than a systematic uncertainty reduction. Sensitivity analysis of the dose reconstruction tells where to improve the state of knowledge so as to reduce dose uncertainty most effectively.

EDUARD HOFER

St. Sebastian Str. 5
84405 *Dorfen, Germany*
elhofer@t-online.de