COMMENTS ON THE OEHHA 2/98 DRAFT HEALTH RISK ASSESSMENT FOR DIESEL EXHAUST

Duncan Thomas
University of Southern California, Department of Preventive Medicine

For presentation at the OEHHA/ARB SRP meeting, ~~February 2, 1998~~ March 18, 1998

I last reviewed the California risk assessment for diesel exhaust in May 1994. It is a formidable document and has, if anything, become more complex since I saw it last. Thus, I've decided to focus my comments today on the issues I raised in my 1994 review and the ways they have been addressed in the current draft. In particular, I will not comment on the extensive review of the animal literature and its use in part of the quantitative risk assessment (QRA), which is beyond my expertise. I will also focus on the carcinogenic effects, specifically lung cancer, that drive the risk analysis.

Let me say at the outset that my overall views have not changed about the carcinogenicity of diesel emissions and the suitability of the epidemiologic data for risk assessment. In 1994, I wrote that I supported the conclusion that the lung cancer effects are causal, and I still find the ensemble of human and experimental evidence convincing in that regard. I also chided the report for failing to include a quantitative meta-analysis of the epidemiologic data (or even explaining why it was not included). That defect has been remedied in this draft with an outstanding meta-analysis, which is in close agreement with one by an independent group that was recently published in *Epidemiology* (Bhatia et al, 1998). Together, these two authoritative meta-analyses support the conclusion that diesel exhausts confer a small but consistent excess risk of lung cancer that cannot be accounted for by smoking or other confounders, thus further supporting my basic conclusion.

In my 1994 review, I also supported the choice of the Garshick et al cohort study as the basis for QRA, although I questioned whether it should be the sole basis, particularly since the individual data that were being touted as a compelling advantage of that study were not used in the QRA except indirectly in later stages. I argued that if that was all that would be done, a similar approach could have been taken with the Garshick case-control data, and perhaps others. I was thus pleased to see the addition of the case-control data to the QRA.

More importantly, however, the entire QRA has been extensively reworked in this draft by extensive reanalyses of the original data and use of appropriate life-table methods to address the concerns I raised about the handling of time-dependent exposures and other time-related factors. It is these issues I wish to focus on in the remainder of the time available. I will try to be brief, in the hopes of simulating discussion that would probably be more rewarding than simply expressing my own views.

There are two main issues we must consider in this QRA. The first is the reanalyses of the original data using various statistical methods to derive a slope coefficient for the dose-response relationship and its confidence limits for use in QRA. The second is the life-table and other techniques used to derive a "unit risk estimate" from these slope estimates.

In the analysis of the cohort data, the over-riding questions are "Is there a positive dose response?" and "Why do the results of the different analyses seem to differ so much?" The report includes three Appendices (D-F) which provide extensive discussion of the reanalyses of these data that were conducted by Dr. Stan Dawson of OEHHA and Dr. Kenny Crump. These, and their summary in section 7.3.4, were hard going for me and are probably incomprehensible to most readers, but the issues raised here are central to the validity of the QRA, so it is worth trying to understand them. The various analyses seem to produce anything from a significant, monotonic, positive dose-response relationship to a significant negative one. So what are we to believe?

Appendices E and F try to lay out the differences between the different methods of analysis and their effects on the findings. Some are trivial, some are major issues. Let me begin by setting aside what I see to be the minor issues: the exclusion of the shop workers, the incomplete follow-up post 1976, the exposure parameter (other than the question of background exposures), the choice of exposure category boundaries, the method of analysis (Cox vs Poisson regression, other than the choice of time scales), and debates over the appropriate use of *p*-values. These are interesting issues that you can read about in the appendices, but in my view, they do not seem to have much influence on the results. The bigger questions are the method of accounting for age and other temporal variables, the treatment of background exposures, and the assumptions made about the pattern of historical exposures. These are all inter-related and subtle.

In my 1984, I argued that age, not calendar-year, was the most important determinant of cancer rates and needed to be carefully controlled in the analysis, either by fine stratification in Poisson regression or by use as the primary time-axis in Cox regression. Furthermore, because calendar year is closely related to cumulative exposure (perfectly so in continuing workers, less so for retirees and those with intermittent exposures), use of calendar year as the primary time-axis in Cox regression could lead to unnecessarily wide confidence limits. These principles are clearly laid out in volume II of Breslow and Day's text. Appendix D provides the results of extensive analyses using a variety of methods to control age, calendar year, and birth cohort effects, which show remarkably little influence on the slope estimates and their standard errors. Generally, I support a flexible modeling strategy in which the "preferred" estimates would be based on the model with the largest Akaike information criterion, which strikes me as providing a reasonable balance between good fit and avoidance of excessive parameters that could spuriously inflate the upper confidence limits. A case could also be made for using the Bayes information criterion, which aims at striking a similar balance, but with respect to minimizing the mean square error of estimation rather than the mean square error of prediction, which I think is more relevant in this context *[check!]*. In every case,

however, the alternative approaches to control of temporal factors produce significantly positive relationships, so we must look elsewhere to account for the negative slopes in earlier analyses. What all the results in tables D-2 and D-3 have in common are the subtraction of background exposures and the use of ramp or roof exposure patterns. It appears from the results in Appendix F that it is only the block exposure pattern that is particularly sensitive to these modeling assumptions, as might be expected from the greater colinearity of that exposure with calendar time.

I must leave it to others to judge the reasonableness of the different exposure patterns, although the rationale for the roof pattern favored by OEHHA seems sensible to me. With that choice, the only remaining major issue seems to be the appropriateness of subtracting background exposures. The report takes the view that the failure to subtract background exposures is the primary reason for the differences between the Crump and Dawson analyses (p 7-28 and F-6), but I find it difficult to understand how this could happen. Some simple mathematics might help elucidate the issues. Suppose we assume a proportional hazards model of the form

$$\lambda[t,Z(t)] = \lambda_0(t) \exp[\beta Z(t)]$$

where Z(t) denotes cumulative exposure to age t, both occupational z(t) and background b(t). Now we can certainly rewrite this model as

$$\lambda[t,Z(t)] = \lambda_0(t) \exp[\beta\{z(t)+b(t)\}] = \lambda^*(t) \exp[\beta z(t)]$$

where $\lambda^*(t) = \lambda_0(t) \exp[\beta b(t)]$ is simply a new "baseline" risk function for the hazard rate for someone exposed only to background levels. Now, the problem arises because we are working with *cumulative* exposures. Occupational exposures z(t) are accumulated since first employment, while background exposures should be accumulated since birth (see Fig F-1). Now, if we simply subtract the cumulative background since first employment, the model is misspecified by a factor $\exp(\beta B t_0)$, where t_0 is age at first employment and B is background concentration. This could lead to a substantial difference between analyses that do or do not adjust for background if there is substantial confounding of cumulative exposure with age at first employment (as seems likely), that might be resolved by including age at first employment as a covariate. I'm not certain that this is the resolution of the apparent paradox, but it bears careful thought.

Before leaving this section on the reanalyses of the original data, I want to comment briefly on the analyses based on the Armitage-Doll multistage model. In my 1994 review, I had specifically suggested that such models be fitted directly to the epidemiologic data, rather than invoked in QRA stage in the rather ad hoc way that had been done. The current draft concludes (p 7-27):

> *...the most accurate models are likely to be the multistage models with a late stage sensitive to diesel exhaust exposure. The unit risks from the multistage mdoels are about 3-fold less than the corresponding general model.*

The crux is whether we have the *right* estimates from the multistage model. Unlike the general multiplicative models used elsewhere, the multistage model makes a very specific prediction for the form of the age dependence of baseline risks (t^{k-1}) and makes no specific allowance for calendar time or birth cohort effects. To be comparable with the results presented elsewhere, one really should allow for the possibility of their confounding effects by allowing a more flexible version of the multistage model with respect to baseline risks, particularly when looking for excess risks that are small in comparison with baseline. I would suggest a multiplicative variant of the model of the form,

$$\lambda[t,Z(t)] = \lambda_0(t)\,[1 + \beta \int z(u)\, g(t,u)\, du]$$

where $g(t,u) = u^{i-1}\,(t-u)^{k-i-1} / t^{k-1}$ and $\lambda_0(t)$ is left unspecified (as in Cox regression) or modeled flexibly in Poisson regression, instead of the more restricted parametric form given at the bottom of p. D-8. It seems plausible to me that the lower unit risk estimates derived from the multistage model might be due in part to misspecification of the baseline dependence on the three time scales (possibly leading to some negative confounding).

A second issue in the multistage models is the choice of stage of action i. The report considers models in which diesel emissions act at the next to the last or the last stage only. The latter seems biologically implausible, as the risk at time t (plus some detection interval) would be determined only by the exposure at that instant and there would be no cumulative effect. A penultimate-stage action (or any earlier stage) does allow a cumulative effect, in this case giving heavier weight to recent exposures and exposures at older ages. This would be the expected pattern for a tumor promotor. For an initiator, one would expect a better fit with i=1 or 2, perhaps, but unfortunately we are not provided the results of any analyses varying the stage of action (other than i=6 and 7). (I note in passing that the multistage model doesn't necessarily assume that exposure affects only a single stage, as stated on p 7-17.) The finding that i=7 fits at all (with a somewhat implausibly long detection interval for lung cancer — 10 years) is due simply to the fact that we do not have detailed exposure data on individuals over time, only these postulated exposure patterns, so that exposures at time t are highly correlated with eariler exposures. Thus, another possible explanation for the lower risks from the multistage model might be or to misspecification of the stage(s) of action (possibly leading to spurious downweighting the effects of earlier exposures). On the other hand, it is probably true that the effect of diesel exposures are strongly modified by age at exposure and latency in ways that are not allowed for in the general multiplicative model, so the use of the multistage model makes *a priori* sense: what these epidemiologic data are probably inadequate for is to estimate the stage of action with any precision.

Finally, let me turn to the lifetable risk assessment methods. In my 1994 review, I extensively criticized the "back of the envelope" calculations based on simple extrapolation of a summary relative risk estimate to the population lifetime lung cancer

risk (with various adjustment factors) and argued that the appropriate way to compute life risk was by a lifetable. I am therefore delighted to see that this advice has been followed, as illustrated in Table 7-8. I have had some difficulty reproducing some of the numbers in that table, which I am currently sorting out with Dr. Dawson, but I want to focus on some more generic questions that affect how this and other risk assessments ought to be done. The *principle* is that the "unit risk" is defined as the increase in lifetime risk of a particular cause of death that is expected to result from a lifetime exposure to a unit concentration of exposure (Thomas et al, *Health Physics* 1992; 63: 259-272). Note my focus on the word "lifetime" *twice* in that definition. The "back of the envelope" calculation approximates this by taking the observed *lifetime* lung cancer risk and multiplying it by the excess relative risk predicted for an *average* lifetime of 70 years. It is only an approximation to the real unit risk when the relative risk varies with age, as it will under any cumulative exposure hypothesis, and will generally tend to underestimate the unit risk because the higher RRs occur when the baseline risk is higher. The lifetable method overcomes this problem by doing a similar calculation at each age, multiplying the age-specific RR (a function of cumulative exposure to that age) by the age-specific baseline rate, and then adding over all ages. But to estimate *lifetime* risk, one must continue the calculations until extinction of the population, not arbitrarily truncate it at age 70. The "unit risks" quoted in the report based on the lifetime method are not lifetime risks at all, but risks to age 70. For a cancer like lung, where most of the deaths (and an even higher proportion of the excess deaths) occur over age 70, this substantially underestimates the unit risks — I estimate by a factor of about 2.5.

In summary, I commend the authors for a fine review of the literature, the addition of a state-of-the-art meta-analysis, for forcefully laying out the case that diesel exhaust is a human carcinogen, and for their extensive reanalyses of the best available epidemiologic data and its incorporation into a much improved risk assessment. As the document has grown, it has necessarily had to address some highly technical, but central, questions and this makes it hard to read. If I had only one wish to make, it would be that someone try to make these sections more accessible to a general readership. Some of the questions that arise are unique to these specific data sets, but many are general and will affect the way such risk assessments are done in the future, both here in California and elsewhere in the world. This risk assessment is being closely scrutinized and could have an impact far beyond the diesel world.