

American Petroleum Institute
1220 L Street, Northwest
Washington, D.C. 20005



Mary Burr Paxton, Ph.D.
Health Scientist
(202) 682-8338
(202) 682-8270 (FAX)

June 17, 1993

Vernon M. Chinchilli, Ph.D.
Center for Biostatistics and Epidemiology
College of Medicine
Pennsylvania State University
Hershey, PA 17033

Dear Vern:

Enclosed at long last is the analysis of the Pliofilm data set by Bob Spirtas and Charlie Brown. The Benzene Task Force would greatly appreciate your reviewing it incisively, on its own merits and with respect to the analysis we did with ENVIRON.

While waiting, I've been exploring why our results differ from Kenny Crump's new results (manuscript enclosed). As is laid out in the table of estimated additional leukemia deaths, the "discrepancies" of concern are:

- ENVIRONS estimated risks are generally lower than Crump's;
- contrary to expectations, for a given analysis, ENVIRON's estimates of risk using the Paustenbach exposure estimates are somewhat higher than those gotten using the Crump and Allen exposure estimates.

As summarized in the next table, I've done a number of additional proportional hazards analyses on the Pliofilm data varying the criteria used to stratify the cohort. The first page (A) is for runs with all 15 cases (i.e., including the female), as used in the ENVIRON analysis; the second page (B) is for runs on the "white males" (non-female, non-black) only, which matches the data set Crump used. The stratification criteria applying to a row are listed in the left-most box. The three main column are for the cumulative exposure estimate used (Rinsky, Crump & Allen, or Paustenbach), each with a column for the slope (β), the standard error of the slope, and its p-value. The possible stratification criteria were:

An equal opportunity employer

BP-00019140

R/S: race and sex (partitioned as black vs. non-black,
female vs. non-female)

loc: plant location at St. Marys or Akron

DOS: date of first exposure, with decades cut at 1940,
etc. for St. Marys and at 1936, etc. for Akron

DOB(o): date of birth, with decades cut offset to match
decades for DOS by location

DOB(s): date of birth, with decades cut the same
regardless of location

A1 are the results in the second ENVIRON paper. B6 would be most comparable to the data set used by Crump, but his model does use age-specific mortality rates. It will be noted (A vs. B) that the inclusion of the female leukemia case and her few stratum mates has virtually no impact on the estimates. If location is omitted as a stratification variable (analyses 4, 5, and 6), the results with the Paustenbach exposures or with the Crump and Allen exposures are very similar.

On page 33 of his manuscript, Crump notes that our dose-response model is exponential and "consequently, although approximately linear at low exposures, exhibits upward curvature." He has asked whether the proportional hazards analysis can be run to make it a linear model directly comparable with his linear models. It can't be, can it?

Crump's second point there is that, because date of first employment (DOS) is likely to be correlated with benzene exposure, stratification on this variable would tend to dampen the measured effect of benzene exposure. The findings for the sets of stratification criteria without DOS would tend to support this assertion in that they are all more significant and have steeper β s than those for A1 reported in our paper no matter which set of exposure estimates is used. Because of the pronounced elevation of the p-values, I'm inclined to believe that the model in #1 is over-parameterized by including both DOS and location. We were endeavoring to mimic Rinsky's matching criteria, but I now think #3 (location without DOS) might have been a better choice. Knowing what we do about the marked qualitative difference between St. Marys and Akron, the stratification on location is desirable, but Crump may have a point in asserting that DOS is too closely related to exposure. The correlations between year of first work and the cumulative exposure measures (see computer printouts) do turn out to be significant, particularly for Akron; interestingly, in St. Marys the cumulative Rinsky exposure estimates do not show such a correlation.

I have run some correlations on the various measures of exposure with each other with wet-side-dry-side and "white male"

partitions, which I have also included for your perusal. As my notations indicated, the perfect explanation of our inverted order of potencies for Crump & Allen vs. Paustenbach exposures did not jump out at me. Perhaps the numbers will say more to you. As I noted in the first of the ENVIRON papers in discussing this apparent inversion (see page 10; page 16, footnote 10; and page 24, Table VII), the mean increase in the workers' exposure estimates with the Paustenbach estimates over those with the Crump and Allen estimates was relatively greater in the non-cases than it was in the cases. This is further documented in the two attached printouts, which show this again and that this had not been the case for the Crump and Allen estimates vs. the Rinsky estimates.

<u>Increase over lower exposure estimates</u>	<u>Non-cases</u>	<u>Cases</u>
Crump and Allen vs. Rinsky	111%	152%
Paustenbach vs Crump and Allen	89%	39%

I've scrutinized the estimates of cumulative exposure and the differences between those obtained using the Paustenbach estimates and those gotten using the Crump and Allen estimates on an individual worker basis. In several hundred instances, the Paustenbach estimate of cumulative exposure is less than the Crump and Allen estimates (as is the case for five of the 15 leukemia cases). In 53 cases, this deficit in the Paustenbach estimate is more than 100 ppm-years; in five more, it exceeds 1000 ppm-years. Referring back to the exposure matrices themselves, it is clear that, although overall the Paustenbach estimates are higher, on a job- and year-specific basis there are numerous exceptions to this trend. As is borne out in practice, there is no reason to expect that an individual's cumulative exposure derived using the Paustenbach exposures will be higher than that derived using the Crump and Allen estimates.

The precise potency estimates gotten by fitting a particular model will depend on the model itself and the exact set of cases and non-cases included in the analysis. Although the basic data set from which ENVIRON and Crump worked were identical, they did not use precisely the same data sets in their actual analyses. All of Crump's analyses employed every "white males." The ENVIRON proportional hazards analyses included the female cases and her stratum mates (although as demonstrated above, the small group of females had virtually no impact on the outcome). More importantly, the exact set of individuals included in a run of our proportional hazards model is determined by the stratification criteria. Because a non-case will contribute information only if he falls in the same stratum with a case who died younger than he did, as the stratification criteria are reduced, two factors work to include more non-cases in the analysis. First, less finely defined strata will contain more non-cases if strata without a case are merged with ones that do have a case. Second, more non-cases in a stratum with a case will be eligible to contribute information if their stratum is

merged with a stratum containing a case who died at a younger age. Increasing the number of individuals included in an analysis will increase its power to generate significant findings.

I next intend to see whether the use of intermediate cumulative exposures (for controls at the age corresponding to a case's death) in the proportional hazards calculations might account for the differences we had from Crump's results using always the lifetime cumulative exposure. Given that most of the cases died at "retirement ages," however, I wonder if this could be the key.

With respect to the inversion in results seen between us and Crump for the Crump and Allen exposure estimates and the Paustenbach exposure estimates, I feel more comfortable that the differences are a function of our models and not an indication that someone has done something wrong.

I'll call you early next week to see what you think we should do about the current version of our the second paper with Risk Analysis. Maybe it would be easier to get together to work out the fine points.

Please keep track of your hours. I got your last bill and entered it into the payment system. We appreciate your efforts.

Sincerely,

Mary Burr Paxton, Ph.D.
Health Scientist

enc: Crump vs. ENVIRON
ph table + comments
Spirtas manu
Crump manu
startyr corr
expo measure corr
diff dist

COMPARISON OF RISK ESTIMATES
FROM
CRUMP AND ENVIRON ANALYSES

Additional leukemia deaths per 1000 workers
with 45 years occupational exposure
to 1 ppm benzene

Model	Exposure Estimates	
	Crump & Allen	Paustenbach
<u>Crump:</u>		
Weighted dose, non-linear outside		
Additive	1.9	0.72
Multiplicative	2.1	0.56
<u>Paxton et al.:</u>		
Proportional hazards stratified on decade of birth, plant location, race, sex		
Decade of starting Pliofilm work	0.26	0.49
Without stratification on start date	0.47	0.64

IMPACT OF STRATIFICATION
ON
PROPORTIONAL HAZARDS ANALYSES

A. Including Female Case

C S r t i r t a e t r i a a	Exposure Estimates Used in Analysis								
	Rinsky			Crump & Allen			Paustenbach		
	β	SE	p-value	β	SE	p-value	β	SE	p-value
1. R/S loc DOB(o) DOS	0.00376	0.00136	0.0023	0.00075	0.00042	0.0602	0.00150	0.00051	0.0015
2. R/S loc DOB(o)	0.00439	0.00122	0.0001	0.00112	0.00038	0.0007	0.00171	0.00046	0.0001
3. R/S loc DOB(s)	0.00512	0.00132	0.0001	0.00136	0.00043	0.0003	0.00199	0.00050	0.0001
4. R/S DOB(s)	0.00488	0.00132	0.0001	0.00146	0.00042	0.0001	0.00172	0.00045	0.0001
5. R/S DOS	0.00441	0.00134	0.0002	0.00160	0.00054	0.0011	0.00165	0.00047	0.0001
6. R/S	0.00521	0.00122	0.0001	0.00167	0.00037	0.0001	0.00177	0.00039	0.0001

IMPACT OF STRATIFICATION
ON
PROPORTIONAL HAZARDS ANALYSES

B. "White Males" Only

C S r t i f i c a t i o n	Exposure Estimates Used in Analysis								
	Rinsky			Crump & Allen			Paustenbach		
	β	SE	p-value	β	SE	p-value	β	SE	p-value
1. R/S loc DOB(o) DOS	0.00376	0.00136	0.0023	0.00075	0.00042	0.0607	0.00148	0.00051	0.0018
2. R/S loc DOB(o)	0.00439	0.00122	0.0001	0.00112	0.00038	0.0007	0.00170	0.00047	0.0001
3. R/S loc DOB(s)	0.00512	0.00132	0.0001	0.00136	0.00043	0.0003	0.00197	0.00050	0.0001
4. R/S DOB(s)	0.00489	0.00132	0.0001	0.00146	0.00043	0.0001	0.00170	0.00046	0.0001
5. R/S DOS	0.00442	0.00134	0.0002	0.00160	0.00054	0.0010	0.00162	0.00047	0.0001
6. R/S	0.00521	0.00122	0.0001	0.00167	0.00037	0.0001	0.00175	0.00039	0.0001