Evaluation of Suitability of Using Fontham et. al. (1994) Study to Derive a Dose Response Model for Environmental Tobacco Smoke

Background

In order to estimate the potential lung cancer risk associated with a specified exposure to environmental smoke (ETS), a mathematical model must be derived. Such a model, which relates the increase in the absolute lung cancer risk to an assumed time dependent ETS exposure is referred to as a dose response model. In order to derive such a model, information demonstrating that lung cancer rates increase as ETS exposure increases must be available. For such information to be useful the observed trend must also be statistically significant (i.e. the probability is small that the observed positive trend was as large as it was without a real ETS effect being present.)

Dr. Steven Bayard of EPA's Office of Health and Environmental Assessment EPA, who is the federal governments primary scientist working on ETS cancer effects analysis, expressed the strong opinion that the best evidence for an lung cancer risk trend is supplied by the Fontham et. al. (1994) study. The single most meaningful part of this study was believed to be the subset of subjects who were exposed to ETS as a child, were self-responders (i.e. were the source of their own exposure estimates to ETS rather than spouse or relative), and the cases were defined as having any type of lung cancer. The specific analysis on this subpopulation regarded as most relevant had the odds ratio adjusted for age, race, study area, fruit consumption, vegetable consumption, supplementary vitamin index, dietary cholesterol and family history of lung cancer. The independent variable was a composite measure of ETS exposure obtained on the job, in the home, and in social settings. The adjusted odds ratio with 95% confidence limits for each of five exposure range categories is displayed below in Table 1. This information is taken from Fontham et. al. (1994)-Table 8.

Table 1- Evidence Considered by EPA to be the Strongest Indication of a ETS Related Trend

Smoke-Years of Exposure During Adulthood	97.5% Upper Bound	Adjusted Odds Ratio	2.5% Lower Bound
None (0)	---	1.0	---
1-11 (6)	5.21	1.85	0.66
12-28 (20)	8.05	2.99	1.11
29-47 (38)	9.00	3.33	1.23
>47 (58)	10.42	3.83	1.41

The purpose of the following analysis is to explore the reasonableness of using the above data set or similar data sets in the Fontham et. al. (1994) study to derive a dose response model for ETS. The first step will be to explore the validity of the claimed positive trends.

Evaluation of Observed Positive Trends

Several factors make it difficult to directly evaluate the validity of the trend test for this data set. First, the complete data set for each individual (except identification information) is required to run an exact statistical test for a trend. Second, a reference to the exact statistical test that was used in obtaining the noted "Trend P=0.001" values was not specified. However, with the given data in the published paper and several reasonable approximations-assumptions, it is possible to evaluate the observed trend in a preliminary manner.

Non Parametric Trend Test

Without making additional assumptions about the information in Table 1, the only test of the null hypothesis that smoke-years of ETS exposure has no effect on the odds ratio that is possible to make is nonparametric in form. We ask the question: if there was no effect of ETS, what is the probability that the five odds ratios were in ascending order of ETS exposure? In total there are 5! = 120 possible orderings of the five odds ratios. Out of these 120 orderings, the single ordering that was most consistent with a monotonically increasing trend was observed. Thus under the null hypothesis, the probability of this occurring by pure chance is $p=1/120=0.00833<0.01$. This suggests that unless, (1) a classification bias whose likelihood or strength or both is proportional to total adult smoke years of ETS exposure exists or, (2) a monotonic association of adult smoke years of ETS with a lung cancer confounder exists, the null hypothesis should be rejected. In other words, the data is consistent with a positive trend due to ETS that is not likely to be the result of random variability.

Evaluation of the Observed Trend in Unadjusted Data

It is recognized that conclusions reached from the analysis of adjusted and unadjusted odds ratio data have the potential for being considerably different. However, it appears from Fontham et. al. (1994) - Table 8 pg. 1758 that the observed crude and adjusted odds ratios are very similar in both size and shape of trend response. Therefore, it is likely that any factor which strongly influences the unadjusted (i.e. crude) analysis would also influence the adjusted analysis. Since sufficient information is available to evaluated the effect of using childhood exposure for the crude case such an analysis will be performed with the anticipation that it would also be relevant for the adjusted case.

If the adjusted odds ratio is plotted against smoke-years in both the exposed and not exposed as a child groups, a clear picture at first ***"appears"*** to emerge. As can be seen in Figure 1, a simple empirical multiple regression model seems to describe the data

relatively well. The interpretation of the model is that both childhood and adult smoke-years have some effect but there also exists a pronounced effect for adult smoke-years-childhood exposure interaction. However, a simple alternative explanation for the observed trends exists that can be evaluated for the unadjusted odds ratios. The cases and controls given in Table 8 of Fontham et. al. (1994) are displayed in Table 2 along with the resulting odds ratio estimates and 95% confidence bounds for: (1) the effect of childhood exposure on adults (i.e. for each exposure group the odds ratio of adult and childhood exposed to only exposed as an adult); (2) the joint total effects childhood and adult exposure (i.e. for each exposure group the odds ratio of adult and childhood exposed to no exposure control group and (3) the effect of adult exposure on individuals also exposed as a child (i.e. for each exposure group the odds ratio of adult and childhood exposed to exposed as a child but not as an adult control group) - Fontham analysis .

Table 2- Alternative Method of Evaluating Effects of ETS Exposure on Odds Ratio for Different Duration and Life Periods of Exposure

	no childhood exp.		childhood exp.				
smoke-years adult exposure	**cases**	**controls**	**cases**	**controls**	**odds ratio childhood exposure (1)**	**odds ratio joint adult and childhood exposure (2)**	**odds ratio effect adult exposure on childhood exposed (Fontham) (3)**
control	23	71	5	44	0.351 (0.098-1.26)	0.351 (0.098-1.26)	1.00
1 - 11	23	90	29	137	0.828 (0.409-1.68)	0.653 (0.319-1.34)	1.86 (0.68-5.10)
12-28	28	97	69	201	1.189 (0.674-2.10)	1.060 (0.615-1.83)	3.02 (1.15-7.93)
29-47	36	97	67	204	0.885 (0.521-1.50)	1.014 (0.588-1.75)	2.89 (1.10-7.59)
≥ 48	31	80	70	182	0.993 (0.566-1.74)	1.187 (0.668-2.05)	3.39 (1.29-8.89)

The results shown in Table 2 are quite surprising. When the data is analyzed in this manner, there is no suggestion of a childhood or a joint childhood and adult effect using approaches (1) and (2) respectively. This is in direct contradiction to the Fontham et. al. (1994) interpretation of a strong effect of adult smoke years when the individuals were exposed as a child which is the last column or method (3). There is no plausible biological explanation for the inconsistency of the above analysis. A simple empirical fact is that for some unknown reason (e.g. random error, bias, etc.) either there are too many lung cancer cases in the no-childhood no-adult exposed control group or too few

lung cancer cases in the childhood exposed-no adult exposure control group. Either way, the interpretation of the data in Table 8 as establishing a positive trend for ETS is not logically justified. The analysis was conducted using only the crude unadjusted rates. However the high level of association noted between the crude and adjusted analyses given in Fontham et. al. (1994)- Table 8 strongly suggests that if access to the complete data set was available, comparable results for an adjusted analysis might also be obtained.

The evidence that childhood exposure has an effect on the odds ratio is non-existent for the unadjusted data. Therefore, it is informative to combine the exposed and not exposed as a child categories to test for a trend in the absence of an assumed childhood exposure effect. Such an analysis is depicted in Table 3.

Table 3- Effect of Adult Smoke Years Independent of Childhood Exposure on Crude Unadjusted Odds Ratio with 95% Confidence Bounds

smoke-years adult exposure	cases	controls	crude unadjusted odds ratio
control	28	115	1.0
1-11	52	227	0.941 (0.290-3.053)
12-28	97	298	1.337 (0.451-3.967)
29-47	103	301	1.405 (0.609-5.298)
≥48	101	262	1.583 (0.814-7.172)

When the artificial effect of dividing the data into exposed and unexposed as a child is removed the apparent trend is not nearly as pronounced as suggested by the Fontham analysis. None of the individual lower confidence bounds for a exposure group is greater than one. An approximate trend test is marginally significant. However, the strength of the ETS effect is estimated to be only about 30% of that obtained by Fontham where the questionable adjustment for childhood ETS exposure is used.

Approximate Linear Trend Test for Adjusted Data

Some preliminary conclusions about the adjusted analysis trend test can be made from the data available in the Fontham et. al. (1994) paper. However, it is apparent that the questionable use of control data noted in the previous section most likely also strongly influences the adjusted results presented by Fontham in Table 8. Even with this potentially strong biasing factor present it is still informative to evaluate the observed trends in the data to see whether a meaningful trend exists under the most favorable interpretation of the data.

It appears that the trend tests employed in the Fontham et. al. (1994) analysis assumed that the bias was zero. In other words, the "regression" line went through the origin on the ln odds ratio scale. If their was a bias this could mean that the observed trend was really an artifact of some extraneous factor not adjusted for in the analysis. This possibility could be tested formally if the data for individuals was available for analysis. Lacking such information, in order to conduct a parametric statistical test it is necessary to make some additional assumptions about the data in Table 1. The first assumption to be made is that the average value in a ETS exposure interval is equal to the midpoint of the closed interval. The second assumption is that the average exposure level in the open interval (i.e. the highest exposure group) is ten smoke-years greater than its lower bound. These are simply reasonable first approximations that are required if further non-parametric approaches to testing trends are to be employed. It would of be far more desirable to have the actual mean values of smoke-years of ETS exposure for each of the exposure groups. In lieu of such information, we shall use as average values the quantities shown in parenthesis ( ) in the first column of Table 1. With this assumed and real information it is possible to conduct approximate trend tests and to investigate the effects of the intercept (i.e. estimated bias) on the significance of the slope (i.e. positive trend).

The form of the relationship that is probably the closest to what is being assumed by the Fontham group is that the increase in the ln of the odds ratio is linearly related to average ETS exposure. Stated in another way, the assumption is that the line passes through the origin, or equivalently, the control value of the ln odds ratio is known to be equal to zero with certainty. This model assumes that the entire increase in the ln odds is due to ETS. To test this hypothesis, the best fitting straight line through the origin is obtained using weighted linear regression where the weights are the inverse of the estimate ln odds variance. The regression line obtained by this technique has the numerical form , ln(odds) = -2.17475+0.028267*(adult smoke-years), where the intercept is fixed at the assumed control ln odds in order to be on a scale where the variance of the observations are independent.

A simple alternative model is that all of the increase in the ln odds is due to a ETS independent bias. This alternative results in a horizontal line that is equal to the weighted mean of the ln odds for the four exposure groups. Other interpretations of the observed

conjunction with a linear trend or a non-linear relationship such as a Michaellis-Menton saturation curve. The goodness of fit of the four mentioned possibilities are displayed visually in Figure 2. It is clear that the relationship most likely used by Fontham, a straight line through the origin (i.e. .....) fits poorly, even worse than assuming all the observed effect is in due to a ETS independent bias (i.e. - - - - - - -) . A simple model that assumes both a combination of constant bias and a linear effect (i.e_____) fits the data adequately and yield a much smaller ETS effect than the straight line through the origin. The fit of the Michaellis-Menton saturation curve (i.e. . _ . _ . _ .) is very close to the observed data and demonstrates that an alternative explanation to what can be interpreted as bias exists. However, it is difficult to interpret the Michaellis-Menton observed relationship within a plausible biological context. The goodness of fit of the four models with appropriate statistical tests are given in Table 4 that confirm the visual impressions.

Table 4 - Goodness of Fit of Various Models to Natural Log Odds
Approximate Adjusted Values Self Responders Fontham et. al. (1994) Table 8

Type of Model	Physical Interpretation of Model	Number of Parameters Estimated	Degrees of Freedom and "p" Value	Weighted Sum of Squares (chi-square)
Weighted Mean	Control and all exposure groups equivalent	1	4 (.0071)	14.072
Linear with intercept fixed at control level	All observed effect is due to ETS (ln odds ratio linear through origin)	2	3 (.0001)	23.136
Control fixed all exposure groups equal to their weighted mean	All observed effect is due to a constant exposure group bias	2	3 (.0370)	8.486
Control fixed linear relationship for exposure groups	Observed effect is due to both bias and ETS exposure	3	2 (.4003)	1.831
Michaellis-Menton saturation with possible bias	Observed ETS effect is sub-linear	3	2 (.9347)	0.135

Conclusions

In the July 14, 1995 issue of Science in a "Special News Report" entitled Epidemiology Faces Its Limits, a number of the world's leading scientists in the field of quantitative epidemiology caution about the use of positive results unless the strength of the association is very strong. They have particular concerns about case control studies which are particularly prone to positive biases. Dr. Norman Breslow, who is very prominent in establishing the mathematical rational upon which case control studies are based, cautions "People (may) think they have been able to control for things that are inherently not controllable." In this light, it is very important not to use an epidemiological study to derive a dose response relationship that does not have unequivocal evidence of a positive trend. At first, it appears that the Fontham et. al. (1994) study may contain very strong evidence of a trend with relative risks in the highest exposure group approaching four. However, two lines of evidence based on summary information contained in the paper suggest that much or all of the observed effect could be due to some unknown systematic bias. First, the shape of the observed dose response appears to be more consistent with all effect due to a bias model than a total effect due to a linear ETS model. The second line of evidence is that when the non significant effect of childhood ETS exposure is ignored in the analysis the trend effect of smoke-years of ETS only marginal significant. This result appears to be primarily due to a deficiency of cases in the exposed as a child but not as an adult control group. Perhaps this might reflect a unconscious bias on the part of lung cancer cases that did not want to blame parents if they were not exposed as adults, so they did not remember childhood exposure. Whatever the true reason for the deficiency of cases (perhaps only random error), the result is biologically inconsistent with any probable sort of ETS effect.

Because of the noted problems with the ETS lung cancer response data, it is highly recommended that a dose response model not be derived from the summary information available in the Fontham et. al. (1994) paper. However this conclusion is based only on partial information. If the detailed information contained in the study was available for analysis an alternative recommendation possibly could be made.