January 2, 2008

CERTIFIED MAIL – RETURN RECEIPT REQUESTED

Document Processing Center (7407M)
(Attn: TSCA 8(e) Coordinator)
Office of Pollution Prevention and Toxics
U.S. Environmental Protection Agency
1200 Pennsylvania Ave., NW
Washington, DC 2046-0001

Contains No CBI

CONTAIN NO CBI

Attn: 8(e) Coordinator

SUBJECT: Quantitative Risk Assessment of Exposure to Butadiene in European Occupational Settings

Shell Chemical LP is submitting the following information under TSCA 8(e), since it may constitute substantial risk information not previously known to the Administrator.

Shell Chemical LP recently received a copy of a study entitled "Quantitative Risk Assessment of Exposures to Butadiene in European Union Occupational Settings Based on the University of Alabama at Birmingham Epidemiological Study: Acute Myelogenous Leukemia, Chronic Lymphocytic Leukemia, and Chronic Myelogenous Leukemia" (copy attached). Although hazard information is in the literature for total and individual types of leukemia, and quantitative risk assessments are available for leukemia (all types combined), this report quantifies risk for three individual leukemia types. This has not been previously reported in the literature.

This report is filed to provide information EPA may find useful. In no way is it intended as a waiver of any rights or privileges belonging to Shell Chemical LP as the reporting corporation, its agents or employees. The reporting corporation, its agents and employees, reserve the right to object to this report's use or admissibility in any subsequent judicial or administrative proceeding against the corporation, its agents or employees.

This report has been compiled based on information available as of the date of filing. The corporation, its agents and employees reserve the right to supplement the data contained in this report, and to revise and amend any conclusions drawn therefrom.

This report contains no Confidential Business Information.

The following person should be contacted if you have questions or a need for discussion:

Michael Hulse
HSSE - SD
Shell Chemical LP
P.O. Box 4320
Houston, TX 77210
Telephone No. 713-241-0032
Facsimile 713-241-3325
email: michael.hulse@shell.com

8EHQ-08-17029

88080000072

308831

Very truly yours,

for

P. J. Snyder, Global Manager
Product Stewardship
Shell Downstream

Attachment

cc: w/attachment
Elizabeth Moran, Lower Olefins Panel Manager
American Chemistry Council
1300 Wilson Blvd.
Arlington, VA 22209

Quantitative Risk Assessment of Exposures to Butadiene in European Union Occupational Settings Based on the University of Alabama at Birmingham Epidemiological Study: Acute Myelogenous Leukemia, Chronic Lymphocytic Leukemia, and Chronic Myelogenous Leukemia

Robert L. Sielken Jr., Ph.D. and Ciriaco Valdez-Flores, Ph.D.
Sielken & Associates Consulting, Inc.
3833 Texas Avenue, Suite 230, Bryan, TX 77802
Tel: 979-846-5175; Fax: 979-846-2671; Email: SielkenAssoc@aol.com

Elizabeth Delzell, Ph.D.
Department of Epidemiology, University of Alabama at Birmingham,
523 Ryals Building, 1665 University Blvd., Birmingham, AL 35294
Dept: 205-934-1200; Tel: 205-934-5857; Email: EDezell2@ms.soph.uab.edu

November 20, 2007

Executive Summary

The American Chemistry Council (ACC) developed a model for the estimation of leukemia risks due to exposures to 1,3-butadiene (BD) based on the University of Alabama at Birmingham (UAB) epidemiological study of workers in the styrene-butadiene rubber industry. The ACC used a statistical-based approach to derive a Poisson regression linear rate ratio model with dose equal to cumulative BD ppm-years that adjusted for variables that had a statistically significant effect on the goodness of fit to the data. The final model derived by the ACC adjusted for age and the number of tasks that involved BD concentrations of 100 ppm or more for any length of time.

The same epidemiological data and methodology used by the ACC in deriving the final model for leukemia mortality is used here to derive models for three subsets of leukemia. The 81 leukemia deaths observed in the UAB data were classified as one of nine mutually exclusive types. Three of those nine types (acute myelogenous leukemia (AML), chronic lymphocytic leukemia (CLL), and chronic myelogenous leukemia (CML)) are well-defined endpoints and include sufficient numbers of decedents in the UAB data to develop models using analyses similar to those presented by the ACC.

The models in this report differ from the ACC model in that the dose metric used here is the cumulative BD ppm-year excluding exposures that are 40 years old or older. In addition to developing Poisson linear rate ratio models similar to those used by the ACC, Cox proportional hazards log linear rate ratio models are also

developed. The Cox proportional hazards log linear model uses individualized data, adjusts for the effect of age in an optimal way, and uses the best estimate of the individual's cumulative BD ppm-years when fitting the data. The Poisson regression model, on the other hand, uses grouped person-years data, adjusts for the effect of age using age groups, and the cumulative BD ppm-years are group values that are dependent on the partitioning used for the fitting. The results based on the more robust Cox proportional hazard analyses are emphasized.

Exposure lags for the three endpoints analyzed herein were determined using a statistically based approach. The fit of the final models for AML and CLL did not improve with lagged cumulative exposures, however, for CMLthe fit of the model improved significantly if cumulative exposures were lagged 15 years.

Cox proportional hazards and Poisson regression models result in the same variables being included as confounders in the final models and both models determine the same lag periods.

Added risks were calculated at an age of 85 years for an occupational exposure to BD for 45 years from ages 20 to 65 years. The calculations were made using SCOEL's methodology, European mortality rates, and European survival probabilities. The maximum likelihood estimate and the 95% upper confidence limit on the added risks of leukemia for an occupation BD exposure of 0.1 ppm based on the ACC final model were 2.91 and 12.03 per million, respectively. Based on the Cox proportional hazards modeling (which is considered by the authors to be most scientifically defensible modeling), the maximum likelihood estimate and the 95% upper confidence limit on the added risks per million for an occupation BD exposure of 0.1 ppm based on the final Cox model for the three endpoints analyzed here are, respectively; 0 and 4.14 for AML, 1.61 and 2.45 for CLL, and 0.07 and 1.88 for CML.

In the final models, the slope of cumulative BD ppm-years in the rate ratio was not statistically significantly different than zero for all leukemia, AML, or CML, but was statistically significantly different than zero for CLL (using Cox proportional hazards modeling but not Poisson regression modeling).

Analyses of the sensitivity of the outcomes to either including or excluding other/unknown and unspecified leukemia types as part of the three endpoints were also performed. The sensitivity analyses show that these other/unknown and unspecified leukemia types hardly affect the outcomes.

Background

In March 2006, a group of experts in different areas of epidemiological research issued a draft report entitled “Cancer Risk Assessment for 1,3-Butadiene (CAS No. 106-99-0)” for the Olefins Panel of the American Chemistry Council[1] (hereinafter referred as the ACC report). This report used the data developed by the University of Alabama at Birmingham (UAB) to derive a dose-response model for leukemia mortality in the Styrene Butadiene Rubber (SBR) worker cohort. A dose-response assessment was developed for leukemia mortality and inhalation exposures to 1,3-butadiene (BD). The assessment reflected the best available science and incorporated the most recent information regarding exposure, epidemiology, toxicokinetics, mode of action, and carcinogenicity of BD. The assessment was prepared with consideration given to the general guidelines for risk assessment as set forth by the National Research Council (1983), and recent guidelines (Chemical and Radiation Leukemogenesis in Humans and Rodents and the Value of Rodent Models for Assessing Risks of Lymphohematopoietic Cancers (USEPA, 1997), Guidelines for Carcinogen Risk Assessment (USEPA, 2005a), and Supplemental Guidance for Assessing Susceptibility from Early-Life Exposure to Carcinogens (USEPA, 2005b)). Several salient points of this work, including the derivation of the model and dose-response analysis, have been published in Sielken *et al.* (2007).

The ACC report focused on the 81 decedents for which leukemia was the cause of death or a contributing cause of death. Poisson regression analyses and the maximum likelihood procedure were used to develop a linear rate ratio model with dose defined as cumulative BD ppm-years (i.e., rate ratio (RR) = 1+ $\beta \times$cumulative BD ppm-years) that best described the observed leukemia mortality experienced by the cohort.

Four non-exposure covariates (age, years since hire, calendar year, and race) and eight exposure covariates (the cumulative number of BD high-intensity tasks or HITs, the cumulative ppm-years of BD exposures above 100 ppm, the cumulative ppm-years of BD exposures below 100 ppm, the cumulative styrene (STY) ppm-years, the number of STY HITs, the cumulative ppm-years of STY exposure above 50 ppm, the cumulative ppm-years of STY exposure below 50 ppm, and the cumulative dimethyldithiocarbamate (DMDTC) mg/cm-years) were evaluated. Age and number of BD HITs were the most statistically significant covariates. With age and the number of BD HITs in the Poisson regression model, the slope of rate ratio model (β) per cumulative BD ppm-year was not statistically significantly different than zero.

[1] Richard Albertini, Elizabeth Delzell, Hong Cheng, Nalini Sathiakumar, Michael Gargas, Lisa Sweeney, Christopher Kirman, Robert Sielken, Ciriaco Valdez-Flores, Jane Teta, and Jeffrey Hicks *Cancer Risk Assessment for 1,3-Butadiene (CAS No. 106-99-0)*, Prepared on behalf of the Olefins Panel, American Chemistry Council, Washington, DC

In the August 28, 2006, report "Quantitative Risk Assessment of Exposures to 1,3-Butadiene in EU Occupational Settings Based on the University of Alabama at Birmingham Epidemiological Study" by Robert L. Sielken Jr. and Ciriaco Valdez-Flores, Ph.D., the final model from the ACC report was coupled with the Scientific Committee Group on Occupational Exposure Limits (SCOEL) methodology for the actuarial computation of added risks for rate ratio models using European Union (EU) population-based age-specific background rates and competing risks. The methodology was applied to an occupational setting corresponding to 45 potential years of exposure to a non-zero concentration of BD. The person was potentially exposed from age 20 to 65 with exposures received more than 40 years previously being ignored. The BD exposure concentration was specified to be the same for each year of employment. The risk calculations were done for an occupational setting with no tasks involving exposures to high BD concentrations (i.e., no tasks involving exposures to concentrations greater than 100 ppm, regardless of duration). The added risk of leukemia mortality from BD exposure was calculated using an age-dependent relative risk model (RR=1 + $\beta \times$ cumulative BD ppm-years, where the cumulative BD ppm-years component is age dependent). The risk coefficient (β) is the slope of the rate ratio that was derived after adjusting for age and the cumulative number of BD HITs (Sielken *et al.,* 2007). Calculations were done for both the maximum likelihood estimate (0.000189) of the slope factor (β) as well as its 95% upper confidence limit (0.000781). The maximum likelihood estimate of the added risk by age 85 is approximately 0.0000029 (0.002910/1,000) when the occupational BD exposure concentration is 0.1 ppm. The 95% upper confidence limit on the added risk by age 85 is approximately 0.000012 (0.012028/1,000).

Follow-up research to the ACC report demonstrated that, although the number of BD HITs and BD ppm-years are correlated, their effects are not entangled in the dose-response modeling; that is, their effects are practically independent. Sample worker exposure profiles suggested that the cumulative BD ppm-years and cumulative number of BD HITs are not necessarily predictive of each other. Furthermore, the cumulative BD ppm-years and the cumulative number of BD HITs can be treated as independent exposure measures in the sense that the methodologies used in the estimation of the time-weighted average (TWA) BD concentrations and the estimation of the number of BD HITs in a shift are independent of each other. In addition, an explicit evaluation of the Poisson regression model versus a more complex model in which BD HITs and BD ppm-years are not disentangled showed that the results of the two models were comparable and that the more complex model did not fit the data statistically significantly better than the model presented by Sielken *et al.* (2007). These results indicate that the Poisson regression model, which treats the effects of BD HITs and cumulative BD ppm-years as independent effects, adequately captured the joint effects of BD HITs and cumulative BD ppm-years. All of these results supported the conclusion that the effects of BD HITs and cumulative BD ppm-years are independent and were successfully disentangled in the Poisson regression model (Sielken *et al.,* 2007).

Introduction

The Lower Olefins Sector Group (LOSG) of the Association of Petrochemicals Producers in Europe (APPE) has sponsored this research with the purpose of extending the risk assessment in the ACC report to include analyses of subsets of leukemia mortality as the cancer endpoints. These new analyses parallel those in the ACC report with additional analyses to include modeling using both the Poisson regression and the Cox proportional hazards methodologies. In addition, a statistically-based lag of cumulative exposures is determined for each endpoint and model. The final models obtained for three subsets of leukemia deaths are further analyzed to characterize the impact of unspecified/unknown types of leukemia on the risk estimates of the subsets of leukemia.

The final models and corresponding sensitivity analyses are then coupled with the SCOEL methodology for the actuarial computation of added risks for rate ratio models using EU population-based age-specific competing risks and background rates for the three subsets of leukemia deaths. The SCOEL methodology is applied to an occupational situation corresponding to 45 potential years of exposure to a non-zero concentration of BD. The worker is assumed to be potentially exposed from age 20 to 65 with exposures received more than 40 years previously being ignored. The BD exposure concentration is specified to be the same for each year of employment. The risk calculations are done for an occupational setting that excludes exposures to covariates and includes the same exposure lags estimated for the final models. The added risks of BD exposures are calculated using an age-dependent relative risk term using both the Poisson regression linear rate ratio model and the Cox proportional hazards log-linear rate ratio model.

Endpoints

The 81 decedents in the UAB epidemiological data set for which leukemia was either the cause of death or a contributing cause of death were subdivided into the following nine mutually exclusive categories:

1 – acute lymphocytic leukemia	(N=3)	ICD9=204.0
2 – acute myelogenous or monocytic leukemia	(N=26)	ICD9=205.0 and 206.0
3 – acute leukemia – other/unknown	(N=4)	ICD9=207.0
4 – chronic lymphocytic leukemia	(N=25)	ICD9=204.1
5 – chronic myelogenous leukemia	(N=16)	ICD9=205.1
6 – chronic leukemia – other/unknown	(N=1)	ICD9=207.1

7 – non-AML – unspecified lymphocytic	(N=2)	ICD9=204.9
8 – non-AML – unspecified myelogenous	(N=3)	ICD9=205.9
9 – non-AML – other non-AML/unknown	(N=1)	ICD9=207.8

The analyses in this report focus on 3 well-defined subgroups of leukemia with adequate numbers of decedents; namely,

AML = acute myelogenous leukemia
CLL = chronic lymphocytic leukemia, and
CML = chronic myelogenous leukemia.

Another well-defined subgroup of leukemia (namely, acute lymphocytic leukemia or ALL) was not considered because there were only three decedents in the UAB cohort for which ALL was the cause or a contributing cause of death.

The initial dose-response modeling has been done with the deaths corresponding to the 3 endpoints (AML, CLL, and CML) defined to be the deaths for workers in subgroups 2, 4, and 5, respectively (that is, the subgroups that are "definite" and include enough number of deaths for quantitative analysis).

In the sensitivity analyses of the dose-response modeling, the final dose-response model for each of the three endpoints (AML, CLL, and CML) is fit to the deaths corresponding to the three endpoints including the "definite" subgroup plus different combinations of the "possible" subgroups. There are two clarifications for the "possible" responses. The first clarification is that, within subgroup 3, the 4 other/unknowns subdivide into 1 "other" (namely, acute erythroleukemia) and 3 "unknowns". In the sensitivity analysis of the dose-response modeling for AML, only the 3 "unknowns" in subgroup 3 are considered for possible inclusion in the set of all workers with deaths corresponding to the AML endpoint. The second clarification is that the one death in subgroup 9 is lymphosarcoma cell leukemia. This leukemia death is not considered for possible inclusion in the subgroups for AML, CLL, or CML.

For the purposes of dose-response modeling, the match-up between the three subgroups of leukemia and the nine categories above is as follows:

Endpoint	Definitely Includes: Primary	Endpoint Possibly Includes: Other/Unknown	Unspecified
AML (N=26 to 32) acute myelogenous leukemia	2 – acute myelogenous leukemia (N=26)	3 – acute leukemia – other/unknown (N=4, see clarification)	8 – non-AML – unspecified myelogenous (N=3)
CLL (N=25 to 28) chronic lymphocytic leukemia	4 – chronic lymphocytic leukemia (N=25)	6 – chronic leukemia – other/unknown (N=1)	7 – non-AML – unspecified lymphocytic (N=2)
CML (N=16 to 20) chronic myelogenous leukemia	5 – chronic myelogenous leukemia (N=16)	6 – chronic leukemia – other/unknown (N=1)	8 – non-AML – unspecified myelogenous (N=3)

The sensitivity analysis of the dose-response modeling includes the uncertainty about which worker deaths with other/unknown/unspecified leukemia subtype correspond to the three endpoints (AML, CLL, and CML).

The excess risk calculations based on the results of the dose-response analyses are for the excess risks of AML, CLL, and CML without the other/unknown/unspecified leukemia subtypes. This means that the population background hazard rates going into the excess risk calculations for AML, CLL, and CML do not include any unspecified/other/unknown form of leukemia.

Table A.1 in appendix A indicates the results of partitioning the 81 workers with leukemia deaths into the nine subsets of leukemia used in this report.

Dose Response Modeling

Poisson regression and Cox proportional hazards analyses are used to evaluate the response data for the three endpoints. As requested by the LOSG, the cumulative exposure to BD (ppm-years) excluding exposures that are 40 years old or older was used as the dose metric in the models relating the rate ratio and exposure to BD. Both Poisson regression modeling and Cox proportional hazard analyses assess the response data grouped by factors (exposure covariates and non-exposure covariates) that may affect the relationship between cumulative BD exposures and the hazard rate of the response. In addition, Poisson regression modeling relies on grouped cumulative exposures to BD and grouped age while the Cox proportional hazard analyses use age as the index variable and the individual's continuous cumulative BD ppm-years exposure in model fitting.

In Poisson regression analysis, the hazard/incidence rate of cancer mortality is modeled as the background rate multiplied by a rate ratio, which is modeled as a function of the individual's age-dependent cumulative dose:

$$\textit{Hazard Rate} = BG \times NECV \times OECV \times F(\textit{BD ppm-years})$$

where,

BG	=	background hazard rate;
NECV	=	non-exposure covariates;
OECV	=	other exposure covariates (other than cumulative BD ppm-years); and
F(BD ppm-years)	=	function of cumulative BD ppm-years

The model used for the Cox proportional hazards analysis has a relationship identical to that of the Poisson regression model but the likelihood function used to evaluate the fit of the model to the observed data is completely different. Appendix B describes the likelihood functions used in the Poisson regression and the Cox proportional hazards analyses.

The Poisson regression analyses that we conducted assume that the relationship between the rate ratio and the cumulative exposure to BD ppm-years is linear. That is,

$$F(\text{BD ppm-years}) = RR = 1 + \beta \times \text{BD ppm-years}$$

The linear model, in addition to being parsimonious and conservative, was chosen to compare with the results for leukemia mortality in the ACC report.

The Cox proportional hazards model, in contrast, assumes that the relationship between the rate ratio and the cumulative exposure to BD ppm-years is log-linear. That is,

$$F(\text{BD ppm-years}) = \log RR = \beta \times \text{BD ppm-years}$$

which is equivalent to,

$$RR = e^{\{\beta \times \text{BD ppm-years}\}}$$

This log linear model was chosen because it is the model closest to the linear model in the low-dose region that could be fitted using the Cox proportional hazards methodology.

The analyses done for this report for each of the three subsets of leukemia (AML, CLL, and CML) parallel those in the ACC report with additional analyses to include modeling using both the Poisson regression and the Cox proportional hazards methodologies. The same statistical approach used in the ACC report is used here to determine the covariates that should be included in the final models.

Poisson regression modeling and Cox proportional hazards analyses are both suitable for adjusting dose-response relationships to incorporate effects of categorical covariates that have an impact on how well the model fits the observed data. The covariates are confounding variables that add noise and bias to the estimated relationship between the response and an exposure variable specified as the explanatory variable in a model (here, cumulative BD ppm-years). There can be two possible errors when considering confounding variables to adjust exposure-response relationships. The first error occurs when there is no attempt to control for a confounder while the second error occurs when a covariate controls for a non-confounder. The first error is potentially more serious than the second error because not adjusting for a confounder may result in a biased dose-response relationship (that is, a biased representation of how the specified explanatory variable affects the response), whereas adjusting for a non-confounder does not usually bias the dose-response relationship but may reduce the precision of the estimates (e.g., p. 94 in Checkoway, et al. 1989).

Although the decision of whether or not to adjust for a confounder should ideally be based on biological or mechanistic information, in this report and in the ACC report, the adjustments for confounders are determined using a statistically based approach. The likelihood ratio test is a standard method used in statistical analysis to determine the impact of adding or subtracting variables to or from a model. The test evaluates the difference between the likelihood of the model with and without the variable being considered for inclusion into the model. The model with the covariate excluded is nested within the model with the covariate included. Therefore, the maximized log likelihood for the model with the covariate included is greater than or equal to the maximized log likelihood for the model with the covariate excluded. The likelihood ratio statistic is simply the negative of twice the difference between the maximized log likelihood for the model excluding the covariate minus the maximized log likelihood for the model with the covariate included (e.g., Breslow and Day, 1980). The distribution of the likelihood ratio statistic is approximately a chi-square distribution with as many degrees of freedom as the number of categories of the covariate minus one. That is,

$$\text{Chi-Square}(n\text{-}1 \text{ degrees of freedom}) = -2 \times \{\text{maximum log likelihood when covariate is excluded} - \text{maximum log likelihood when covariate is included}\}$$

where n is the number of categories for the covariate considered for inclusion in the model. Because there are n-1 coefficients that are either estimated or set to zero and the n coefficients are normalized, there are n-1 degrees of freedom.

In the analyses there are five non-exposure covariates;

a. age,
b. years since hire,
c. calendar year,
d. race, and
e. plant;

and six exposure covariates;

a. cumulative exposure to styrene (STY) ppm-years,
b. cumulative exposure to dimethyldithiocarbamate (DMDTC),
c. cumulative number of BD HITs,
d. cumulative number of STY HITs,
e. cumulative exposure to STY ppm-years to concentrations below 50 ppm, and
f. cumulative exposure to STY ppm-years to concentrations above 50 ppm),

that are considered for inclusion in the model. For each primary subset of leukemia deaths (AML, CLL, CML), five categories for each continuous non-exposure covariate (age, years since hire, and calendar year) were chosen so that there was approximately the same number of the leukemia-type specific decedents in each of the five categories. Similarly, one category for non-exposed person-years and five categories for exposed person-years for each subset of leukemia deaths (AML, CLL, CML) of the six exposure covariates were chosen so that there was approximately the same number of the leukemia-type specific decedents in each of the five exposure categories. Race was grouped into black and non-black workers while plant was grouped into the six plants for which there was exposure information. These eleven covariates were chosen a priori based on previous risk assessments. Here, however, the eleven covariates are considered for inclusion in the model in a sequential way using the likelihood ratio test to determine whether the contribution of each covariate to the maximized log likelihood is statistically significant.

Although the Poisson regression analyses adjust for age as a covariate effect, the Cox proportional hazards methodology intrinsically adjusts for age by using age as the variable indicating when events occur and when the values of covariates are calculated. Thus, when considering covariates for inclusion in the model, the models based on the Poisson regression methodology explicitly consider the effect of age as a covariate effect but the models based on the Cox proportional hazards model already adjust for the effect of age and do not explicitly considers age as a covariate to be included in the model .

The person years in the UAB cohort were partitioned into eleven different categories of cumulative BD exposure for each of the three subsets of leukemia. The first category corresponds to the person years with zero BD exposure. The other ten categories correspond to deciles in the distribution of cumulative ppm-years of each of the three primary subsets (AML, CLL, and CML) of leukemia decedents that were exposed to BD. Cumulative BD exposure was modeled as a continuous predictor variable with the average cumulative exposure (averaged over the person years in the category) as the reference point for modeling purposes. For models based on the Cox proportional hazards methodology, however, the grouping of cumulative BD exposures was not necessary because this methodology uses each individual's cumulative BD exposure.

Cumulative BD ppm-years excluding exposures that occurred 40 or more years ago is used as the explanatory variable and the hazard rates of the three subtypes of leukemia are the response (dependent) variables in the Poisson regression and the Cox proportional hazards modeling. The analysis to determine which covariates should be part of the model considers both non-exposure covariates and exposure covariates simultaneously. Covariates are sequentially added to the model based on the maximum statistically significant improvement in the likelihood.

Thus, the analysis of the three endpoints (AML, CLL, CML) starts with a fixed partition of the person years (here, partitioned into 11 intervals according to cumulative BD ppm-years) for the Poisson model and the continuous BD ppm-years for the Cox proportional hazards model. Then the procedure determines whether the model with no covariates or explanatory variables is improved if the model for the rate ratio includes cumulative BD ppm-years as the explanatory variable. The next step in the sequential analysis is to determine what categorical covariates further improve the likelihood of the rate ratio model fit. The analyses are done separately and independently for each of the three subsets of leukemia (AML, CLL, and CML) and each of the two modeling approaches (Poisson regression and Cox proportional hazards).

Appendix C shows details of the statistical procedure used to determine the model and covariates that best fit the observed hazard rates of each of the three subsets of leukemia for the Poisson regression linear rate ratio models and the Cox proportional hazards log linear rate ratio models. Table 1 shows the maximum likelihood estimates and the standard error of the estimates for each of the three subsets of leukemia deaths and each of the two modeling techniques. The table also shows the covariates included to adjust the dose-response relationship because they have statistically significant impacts on the likelihoods of the models fit to the data. The results in Table 1 do not include any exposure lags other than the exclusion of exposure more than 40 years ago.

Table 1. Maximum likelihood estimates and standard errors of the estimates for the three subsets of leukemia deaths for the models that adjusted for the statistically significant covariate effects using the Poisson regression linear rate ratio model and the Cox proportional hazards model (models without lags)

Endpoint: subset of leukemia	Poisson regression linear rate ratio model		Cox proportional hazards log linear rate ratio model	
	MLE (S.E.)	Statistically significant covariates included	MLE (S.E)	Statistically significant covariates included
AML	-4.75×10^{-4} (4.35×10^{-4})	Age, Cumulative DMDTC exposure	-2.20×10^{-4} (5.67×10^{-4})	Age, Cumulative DMDTC exposure
CLL	2.85×10^{-3} (2.08×10^{-3})	Age	$4.15\times10^{-4**}$ (1.32×10^{-4})	Age
CML	1.28×10^{-4} (6.39×10^{-4})	Age[1], Number of BD HITs	-1.29×10^{-4} (6.08×10^{-4})	Age, Number of BD HITs

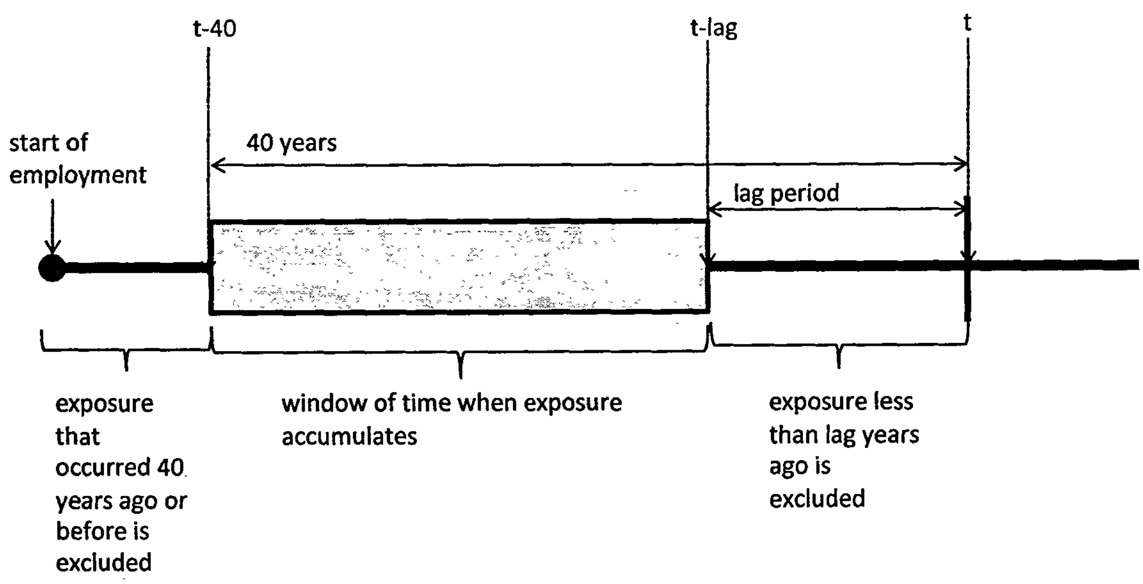
[1]for CML and the Poisson linear rate ratio model the statistically significant covariates were years since hire and number of BD HITs, but age and number of BD HITs were used instead for reasons outlined in Appendix C
**statistically significant at the 1% significance level using Wald's test

The Poisson regression linear model and the Cox proportional hazards log linear model result in identical statistical conclusions regarding what covariates should be included in the model for each of the three endpoints considered. All of the slopes for cumulative BD ppm-years in the Poisson regression linear rate ratio model that adjust for the statistically significant covariates are not different from zero at the 5% significance level. The same is true for the Cox proportional hazards log linear model except for the coefficient of cumulative BD ppm-years in the model of CLL rate ratios which is statistically significantly greater than zero at the 1% level. The coefficient for the Cox log linear model of CLL is approximately 7-fold less than the slope of the Poisson linear rate ratio. The reliance of the Poisson regression procedure on grouped data can exaggerate the slope if it assigns the average of cumulative BD ppm-years in the highest exposure group to all observations in that category. The Cox proportional hazards model, on the other hand, uses each individual's cumulative BD ppm-years, as opposed to grouped average exposures, and results in estimates of the dose-response relationship that are more consistent with the observed cancer mortality experience.

Lags in Cumulative Exposures

Lagging exposures to BD has been considered in previous risk assessments. For example, the EPA's Science Advisory Board (1998) recommended that, in addition to exploring the importance of the number of BD HITs, the effects of the dose-time relationship be explored. The SAB suggested considering "a model that assumes a limited effect time (i.e., that leukemia risk during a given year of age is affected largely by the butadiene exposures received during the previous, say, 20 years, and only slightly or not at all by more distant ones)." As requested by the LOSG the cumulative BD ppm-years and other exposure covariates are calculated excluding exposures that occurred 40 or more years ago. Here, the effects of additionally excluding the most recent years of exposure (e.g., exposures within the last five, 10, ..., 35 years from the observation time) on the likelihood of the model fit to the data are explored. This period of time of recent exposures that are not included in the calculation of cumulative exposures is called the lag period. The cumulative exposure at any time t is the sum of exposures that occurred between 40 years ago and lag years ago. Figure 1 shows a graph of how the lagged cumulative exposure at time t is calculated for an individual when exposures that occurred 40 or more years ago are also excluded.

Figure 1. Schematic representation of lagged cumulative exposure that occurred between 40 years ago and lag years ago



The models that included the statistically significant covariates (derived including all exposures that occurred less than 40 years ago) are used as the starting point to do a likelihood based search for the lag period that maximizes the likelihood of the observed mortality of AML, CLL and CML. Lagging exposures for five, 10, 15, 20, 25, 30 and 35 years have been considered in these analyses. As can be inferred from Figure 1, lagging exposures 40 or more years would result in zero cumulative exposures.

When exposures during a period of time are excluded (whether it is recent years, distant years, or both being excluded), the values of all exposure variables are changed. For example, both cumulative BD ppm-years and cumulative number of BD HITs are reduced. This change in cumulative exposures affects the both the number of deaths with the endpoint and the number of person years assigned to each group of cumulative exposures in the Poisson regression models. Thus, the standard Poisson regression likelihood, which is based on the number of responses and person-years in each group, is not robust to changes in the lag period. The Cox proportional hazards model, on the other hand, is robust to changes in the values of variables. That is, the likelihood function is well behaved and comparable to the likelihood function with a different set of values of the lag period. This is primarily due to the fact that the Cox proportional hazards model calculates the likelihood based on the individual's observation time and those observation times are the same regardless of how the exposures are evaluated.

The standard Poisson regression analysis relies on grouped person-years to calculate the likelihood. These groups of person years change if the cumulative exposures change due to the way they are evaluated causing instability (lack of comparability) in the likelihood of the standard Poisson analysis. To alleviate this shortcoming of the standard Poisson regression likelihood function in the search for an optimal lag period, an individualized Poisson regression likelihood function has been used. The individualized Poisson regression likelihood, instead of being based on groups of person years, is based on the individual's likelihood of the response given the partition of the person years. This individualized Poisson regression likelihood is more stable than the standard Poisson regression likelihood when there are changes in cumulative exposures due to different lag periods. Appendix B discusses the likelihood functions for the Cox proportional hazards model, the standard Poisson regression analysis, and the individualized Poisson regression likelihood evaluation.

Although the period of exposure that is relevant for each endpoint should ideally be based on mechanistic information (as may have been the case with the decision of excluding exposures that occurred 40 years ago or before), in this report the exposure lag is determined using a statistically based approach. The likelihood ratio test is a standard method used in statistical analysis to determine the impact of adding exposures lags to the model. The test evaluates the difference between the likelihood of the model allowing lagged exposures and the likelihood of the model with non-lagged exposures. The model with non-lagged exposures is nested within the model allowing lagged exposures. Therefore, the maximized log likelihood for the model allowing exposures lags is greater than or equal to the maximized log likelihood for the model with non-lagged exposures. The likelihood ratio statistic is simply the negative of twice the difference between the maximized log likelihood for the model with non-lagged exposures minus the maximized log likelihood for the model allowing lagged exposures (e.g., Breslow and Day, 1980). The distribution of the likelihood ratio statistic is approximately a chi-square distribution with one degree of freedom for the one value (lag) that is allowed to vary. That is,

Chi-Square(1 degrees of freedom) = - 2 ×
{ maximum log likelihood with non-lagged exposures -
maximum log likelihood with lagged exposures }

The analyses of the three endpoints (AML, CLL, and CML) and two modeling approaches (Poisson regression and Cox proportional hazards) start with the model (adjusting for statistically significant covariates) that best fit the observed mortality for each endpoint. The procedure determines whether a lag improves the fit of the model by comparing the maximum likelihood value with the model with the lagged exposure with the maximum likelihood value of the model with the non-lagged exposure. This comparison is made for each of the lag periods considered. A non-zero lag period is selected if the model fits the data statistically significantly better than the model with non-lagged exposure and if the maximum likelihood value of the model with this non-zero lag is the largest among all the maximum likelihood values of the model with non-zero lags.

Appendix D shows details of the statistical procedure used to determine the exposures lags that best fit the observed hazard rates of each of the three subsets of leukemia for the Poisson regression linear rate ratio models and the Cox proportional hazards log linear rate ratio models. Table 2 below shows the maximum likelihood estimates and the standard error of the estimates for each of the three subsets of leukemia deaths and each of the two modeling techniques using the exposure lags that significantly improved the maximized likelihood of non-lagged exposures.

Table 2. Maximum likelihood estimates and standard errors of the estimates for the three subsets of leukemia deaths for the models that adjusted for the statistically significant covariate effects and included statistically significant exposures lags using the Poisson regression linear rate ratio model and the Cox proportional hazards model

Endpoint: subset of leukemia	Poisson regression linear rate ratio model			Cox proportional hazards log linear rate ratio model		
	MLE (S.E.)	Statistically significant		MLE (S.E)	Statistically significant	
		covariates included	Lag (years)		covariates included	Lag (years)
AML	-4.75×10^{-4} (4.35×10^{-4})	Age, Cumulative DMDTC exposure	0	-2.20×10^{-4} (5.67×10^{-4})	Age, Cumulative DMDTC exposure	0
CLL	2.85×10^{-3} (2.08×10^{-3})	Age	0	4.15×10^{-4}** (1.32×10^{-4})	Age	0
CML	-2.26×10^{-4}[!] (5.51×10^{-4})	Age[1], Number of BD HITs	15	4.11×10^{-5} (7.22×10^{-4})	Age, Number of BD HITs	15

[1]for CML and the Poisson linear rate ratio model the statistically significant covariates were years since hire and number of BD HITs, but age and number of BD HITs were used instead for reasons outlined in Appendix C

**statistically significant at the 1% significance level using Wald's test

[!]in order to be consistent with other slope estimates, this estimate is based on the standard maximum likelihood procedure for the Poisson regression model which is slightly different than the estimate based on the individualized maximum likelihood procedure used in determining the lag itself

Models from both, the Poisson regression and the Cox proportional hazards methodologies, result in the same conclusions for the length of the exposure lags for the three subsets of leukemia. There were no exposures lags that significantly improved the likelihood of the fit for AML and CLL. For CML, however, a 15-year lag results in statistically significant improvements in the maximum likelihood fits over the models with non-lagged exposures for both the Poisson regression model and the Cox proportional hazards model. The best estimate of the slope for the Poisson regression model and CML went from non-significant positive (1.28×10^{-4}) for non-lagged exposures to non-significant negative (-2.26×10^{-4}) for exposures lagged 15 years. The opposite occurred for the Cox proportional hazards model and CML; that is, the slope estimate of the log linear model for non-lagged exposures went from non-significant negative (-1.29×10^{-4}) to non-significant positive (4.11×10^{-5}) for exposures lagged 15 years.

Unknown/Unspecified Leukemia Types

The final models that adjust for statistically significant covariate effects and include any statistically significant lag period were all based on the three subsets of leukemia that could be definitely classified as AML, CLL, or CML. The sensitivity of the model slopes to the uncertainty associated with the leukemia types that could not be definitely assigned to one of the three primary subsets of leukemia is presented here. In this sensitivity analysis of the dose-response modeling, the final dose-response model for each of the three endpoints is fit to the deaths corresponding to the three endpoints including the "definite" subgroup plus different combinations of the "possible" subgroups. The sensitivity analysis of the dose-response-modeling includes the uncertainty about which worker deaths with unspecified/unknown leukemia subtype correspond to the three endpoints (AML, CLL, and CML).

Tables 3, 4 and 5 list the results of the sensitivity analyses for each of the AML, CLL and CML, respectively, and for the Poisson regression model and the Cox proportional hazards model.

Table 3. Sensitivity of the final model for acute myelogenous leukemia (AML) to including workers deaths with other/unknown acute leukemia and/or unspecified myelogenous leukemia

Acute Myelogenous Leukemia							
End Point	ICD Revision 9				Cases	Poisson Regression (Covariates: Age, DMDTC) (Lag: 0 years) MLE (S.E.)	Cox Proportional Hazards (Covariates: Age, DMDTC) (Lag: 0 years) MLE (S.E.)
	205.0	206.0	205.9	207.0			
AML	X	X			26	-4.75×10^{-4} (4.35×10^{-4})	-2.20×10^{-4} (5.67×10^{-4})
AML + acute leukemia (other/unknown)	X	X		X	29	-2.86×10^{-4} (5.03×10^{-4})	-1.47×10^{-4} (4.53×10^{-4})
AML + Non-AML (unspecified myelogenous)	X	X	X		29	-1.96×10^{-4} (5.61×10^{-4})	-9.71×10^{-5} (4.53×10^{-4})
AML + acute leukemia (other/unknown) + Non-AML (unspecified myelogenous)	X	X	X	X	32	-3.34×10^{-5} (6.06×10^{-4})	-6.53×10^{-5} (3.83×10^{-4})

The results in Table 3 show that for AML the maximum likelihood estimates of the models are non-significant negative slopes, regardless of which unknown/unspecified leukemia types are included with AML. That is, the data does not support an increasing hazard rate of AML with increasing exposure to

BD ppm-years. This result applies to both, the Poisson regression linear rate ratio and the Cox proportional hazards log linear rate ratio models adjusted for age and cumulative exposure to DMDTC.

Table 4. Sensitivity of the final model for chronic lymphocytic leukemia (CLL) to including workers deaths with other/unknown chronic leukemia and/or unspecified lymphocytic leukemia

Chronic Lymphocytic Leukemia						
End Point	ICD Revision 9			Cases	Poisson Regression (Covariates: Age) (Lag: 0 years) MLE (S.E.)	Cox Proportional Hazards (Covariates: Age) (Lag: 0 years) MLE (S.E.)
	204.1	204.9	207.1			
CLL	X			25	2.85×10^{-3} (2.08×10^{-3})	$4.15 \times 10^{-4**}$ (1.32×10^{-4})
CLL + chronic leukemia (other/unknown)	X		X	26	3.34×10^{-3} (2.27×10^{-3})	$4.12 \times 10^{-4**}$ (1.28×10^{-4})
CLL + Non-AML (unspecified lymphocytic)	X	X		27	2.19×10^{-3} (1.66×10^{-3})	$4.01 \times 10^{-4**}$ (1.33×10^{-4})
CLL + chronic leukemia (other/unknown) + Non-AML (unspecified lymphocytic)	X	X	X	28	2.60×10^{-3} (1.81×10^{-3})	$3.98 \times 10^{-4**}$ (1.30×10^{-4})

** statistically significantly greater than zero at the 1% significance level

The results in Table 4 show that for CLL the maximum likelihood estimates of the Poisson regression linear rate ratio models are non-significant positive slopes, regardless of which unknown/unspecified leukemia types are included with CLL. The best estimates and standard errors of the slope of the Poisson linear rate ratios are within 50% of each other.

In Table 4 also, the slopes for the Cox proportional hazards log linear rate ratio model are statistically significantly greater than zero for all combinations of possible leukemia types that could be classified as CLL. The maximum likelihood estimates and standard errors for the Cox proportional hazards model are within five percent of each other.

The slopes of the Cox proportional hazards log linear model, besides being statistically significantly greater than zero, are more than half an order of magnitude smaller than the non-significant slopes of the Poisson linear model for CLL. The larger slope estimates of the Poisson linear rate ratio model are caused primarily by the reliance of this approach on the average of the person-years grouped by categorization of the cumulative exposures to BD instead of using the individual's cumulative exposure to BD. The Cox proportional hazards model, on the other hand, uses the individual's cumulative BD exposure thus avoiding the shortcomings of the Poisson regression model with grouped person years. The differences in the slope estimates are also caused by the differences in the form of the model, the differences in the adjustment for age, and the functional differences of the likelihood evaluation.

Table 5. Sensitivity of the final model for chronic myelogenous leukemia (CML) to including workers deaths with other/unknown chronic leukemia and/or unspecified myelogenous leukemia

Chronic Myelogenous Leukemia						
End Point	ICD Revision 9			Cases	Poisson Regression (Covariates: Age, BD HITs) (Lag: 15 years) MLE (S.E.)	Cox Proportional Hazards (Covariates: Age, BD HITs) (Lag: 15 years) MLE (S.E.)
	205.1	205.9	207.1			
CML	X			16	-2.26×10^{-4} (5.51×10^{-4})	4.11×10^{-5} (7.22×10^{-4})
CML + chronic leukemia (other/unknown)	X		X	17	-2.07×10^{-4} (5.79×10^{-4})	9.80×10^{-5} (6.21×10^{-4})
CML + Non-AML (unspecified myelogenous)	X	X		19	1.75×10^{-4} (8.10×10^{-4})	8.99×10^{-5} (6.05×10^{-4})
CML + chronic leukemia (other/unknown) + Non-AML (unspecified myelogenous)	X	X	X	20	2.15×10^{-4} (8.60×10^{-4})	1.33×10^{-4} (5.41×10^{-4})

The results in Table 5 show that for CML the maximum likelihood estimates of the Poisson regression linear rate ratio models and the Cox proportional hazards models are values non-statistically significantly different from zero slopes, regardless of which unknown/unspecified leukemia types are included with CML.

The best estimates of the slope of the Poisson linear rate ratios change sign depending on the unknown/unspecified leukemia type assigned to CML. This change of sign in the best estimates, however, is mostly irrelevant because both the negative slopes and the positive slopes are not statistically significantly different from zero.

For CML, both the Poisson regression and the Cox proportional hazards models with exposures lagged 15 years and adjusted for age and the number of BD HITs results in non-significant dose-response relationships. Thus, the data does not support an increasing hazard rate of CML with increasing exposure to BD ppm-years regardless of which unknown/unspecified leukemia types are assigned to CML.

Risk Characterization

The final Poisson and Cox models with their statistically significant covariates and statistically significant lag periods are used to estimate added cancer risks for AML, CLL and CML. European-specific age-adjusted background mortality rates for AML, CLL and CML along with age-adjusted survival probabilities are used to calculate added risks using SCOEL's methodology. The lifetime added risks are calculated for non-zero occupational exposures to BD during 45 potential years of work with the persons being exposed between 20 and 65 years of age and ignoring exposures received 40 or more years ago and also excluding exposures received within the lag period. It is assumed that there are no tasks involving exposure to BD concentrations higher than 100 ppm (i.e., no BD HITs) and that there are no other exposures that could contribute to the development of AML, CLL or CML. The occupational BD exposure concentration is specified to be the same for each year. The occupational BD exposure concentration is specified to be either 0.1, 0.2, 0.5, 1, 2, 5, or 10 ppm. The added risks calculated using SCOEL methodology are approximately proportional to the occupational exposure concentrations between 0.1 and 10 ppm for the Poisson linear rate ratio model and between 0.1 and 2 ppm for the Cox proportional hazards log linear model. Taking advantage of this proportionality of added risks and occupational BD concentrations, the added risk at one concentration level is discussed here, and the conclusions are extended to other exposure concentrations. In this section, the results for the added risks for an occupational exposure to a BD concentration of 0.1 ppm are used for discussion purposes while the specific results for all the occupational concentrations 0.1 to 10 ppm are given in Appendix E.

In Table 6 the added risks (number of additional deaths with the response per 1,000 individuals) in an 85-year lifetime are given for the three endpoints (AML, CLL, and CML) and the maximum likelihood estimates of the slopes in the Poisson and Cox models.

Table 6. Lifetime added risks per 1,000 individuals for an occupational BD exposure[1] concentration of 0.1 ppm using the maximum likelihood estimate of the Poisson regression linear model and the Cox proportional hazards log linear model for three primary and three alternative definitions of subsets of leukemia (AML, CLL, and CML)

Leukemia Primary Subset	Model	Leukemia subtypes included with the primary subset			
		Primary	Primary + Other/Unknown	Primary + Unspecified	Primary + Unspecified + Other/Unknown
AML	Poisson	0[2]	0	0	0
	Cox	0	0	0	0
CLL	Poisson	0.01104	0.01293	0.00897	0.01065
	Cox	0.00161	0.00160	0.00164	0.00163
CML	Poisson	0	0	0.00033	0.00041
	Cox	0.00007	0.00016	0.00017	0.00025

[1]45 potential years of work with the persons being exposed between 20 and 65 years of age and ignoring exposures received 40 or more years ago and also excluding exposures received within the lag period
[2]an added risk equal to 0 indicates that the slope of the model was non-positive, resulting in a dose-response relationship that would not estimate a positive added risk for any positive exposure to BD

The maximum likelihood estimates of the slopes of the Poisson linear rate ratio and of the Cox proportional hazards log-linear models are less than zero for AML and any other unspecified/other/unknown leukemia that could be classified as AML. Consequently there is no added risk for this AML in Table 6.

The added risks of CLL calculated with the maximum likelihood estimates of the Poisson regression model are between 5.5- and 8-fold larger than the ones calculated with the Cox proportional hazards model. This is consistent with the non-significant estimates of the Poisson linear model being more than half an order of magnitude greater than the significant slopes of the Cox proportional hazards log linear model for CLL. The larger slope estimates of the Poisson linear rate ratio model, as indicated above, are caused primarily by the grouping of person-years by categories of cumulative exposures to BD, the differences in the form of the models, the differences in the adjustment for age, and the functional differences of the likelihood evaluation. Although the added risks calculated for CLL are smaller with the Cox proportional hazards model in Table 6, they are based on estimates of slopes that are statistically significantly greater than zero whereas the slopes for the Poisson regression model are not statistically significantly greater than zero.

The classification of unknown and unspecified leukemia types as CLL has little impact on the calculated added risk for both the Poisson model and the Cox model. The added risks of CLL in Table 6 for the Poisson model vary from 8.97

to 12.93 in a million and from 1.60 to 1.64 in a million for the Cox model.

The maximum likelihood estimates of the slopes for both Poisson and Cox models were not statistically significantly different from zero for CML. The estimate of the slope for the Poisson linear model was less than zero for the definite CML and for the combined definite CML and other/unknown types while it was greater than zero for the combined definite CML and unspecified types and for the combined definite CML, other/unknown and unspecified types. In Table 6, the positive added risks based on the Poisson model are less than a factor of two greater than the positive added risks based on the Cox model. The added risks calculated with the definite CML are smaller than the added risks calculated with other/unknown and unspecified types of leukemia that could be counted as CML.

Table 6 shows the results using the maximum likelihood estimates of the slopes for the Poisson and Cox models. The calculated added risks are zero for AML, are larger for CML and even larger for CLL. The positive added risks for CLL are between 6- and 27-fold larger than the positive added risks for CML based on the same model and group of other and unknown leukemia types.

In Table 7 the added risks (number of additional deaths with the response per 1,000 individuals) at an 85-year lifetime are given for the three endpoints (AML, CLL and CML) and the 95% upper confidence limits on the of the slopes in the Poisson and Cox models.

Table 7. Upper bounds on the lifetime added risks per 1,000 individuals for an occupational BD exposure[1] concentration of 0.1 ppm calculated using the 95% upper confidence limit on the estimated slopes of the Poisson regression linear model and the Cox proportional hazards log linear model for three primary and three alternative definitions of subsets of leukemia (AML, CLL, and CML)

Leukemia Primary Subset	Model	Leukemia subtypes included with the primary subset			
		Primary	Primary + Other/Unknown	Primary + Unspecified	Primary + Unspecified + Other/Unknown
AML	Poisson	0.00140	0.00329	0.00445	0.00616
	Cox	0.00414	0.00364	0.00397	0.00361
CLL	Poisson	0.02429	0.02739	0.02016	0.02285
	Cox	0.00245	0.00241	0.00254	0.00251
CML	Poisson	0.00114	0.00125	0.00288	0.00311
	Cox	0.00206	0.00188	0.00207	0.00195

[1]45 potential years of work with the persons being exposed between 20 and 65 years of age and ignoring exposures received 40 or more years ago and also excluding exposures received within the lag period

Although the maximum likelihood estimates of the slopes for the Poisson and

Cox models were negative for AML, the 95% upper confidence limits on the estimates of the slopes are positive. The Poisson model was more sensitive than the Cox model to the uncertainty of what unknown or unspecified leukemia types were considered AML. The 95% upper confidence limit on the added risks for AML in Table 7 are between 1.40 and 6.12 in a million (i.e., a factor of 4.4 between the minimum and the maximum) for the Poisson model and between 3.61 and 4.14 in a million (i.e., a factor of 1.1) for the Cox model.

The 95% upper confidence limits on the added risks for CLL in Table 7 calculated with the Poisson models are between 8- and 10-fold greater than the risks calculated with the Cox model. The effect of the classification of unknown and unspecified leukemia types as CLL is negligible in the calculated 95% upper confidence limits on added risk for both the Poisson model and the Cox model. The risks of CLL in Table 7 for the Poisson model vary from 20.16 to 27.39 in a million (i.e., a factor of 1.4) and from 2.41 to 2.54 in a million (i.e., a factor of 1.1) for the Cox model.

For CML, the calculated added risks based on the 95% upper confidence limits on the slopes of the Poisson models are more erratic than those based on the 95% upper confidence limits on the slopes the Cox models. The added risks go from 1.14 to 3.11 per million for the Poisson models (a factor of 2.7) and from 1.88 to 2.07 (a factor of 1.1) for the Cox models for different subtypes included with CML. The added risks with the Poisson and the Cox models are, however, within a factor of 2 from each other for the same definition of CML.

Table 7 shows the results using the 95% upper confidence limits on the estimates of the slopes for the Poisson and Cox models. For the Cox models, the smallest added risks are for CML, the largest are for AML and the added risks CLL are somewhere in-between. The range, however, between the smallest and the largest of the twelve added risks for the 95% upper confidence limits on the slopes of the Cox models is only 2.2 (i.e.,0.00414 for AML primary / 0.00188 for CML primary plus other/unknown).

The upper bounds on the added risks based on the 95% upper confidence limits on the estimated slopes of the Poisson model are very different depending on the subset of leukemia. The lowest added risks are for CML, followed by AML and with the largest added risks being for CLL. The smallest added risk for CLL (0.02016) is more than 3-fold greater than the largest added risk for AML or CML (0.00616). The calculated added risks range from 0.00114 to 0.02739 (a 24-fold difference) for the Poisson regression model.

While all but one of the calculated added risks based on the 95% upper confidence limits on the slopes of the Poisson and Cox models are within a factor of two from each other for AML and CML, for the CLL group the added risks for the Poisson models are more than a factor of 7.9 greater than the added risks for the Cox models.

Table 8 shows the ratio of the added risks shown in Tables 6 and 7. The ratio of the added risk of CLL based on the 95% upper confidence level to the maximum likelihood estimate of the slope for the Poisson model is approximately 2.2 and is only 1.5 for the Cox model. The uncertainties on the estimated slopes for CML are much greater, resulting in ratios between 7.7 and 29.9 for the Cox models and at least 7.6 for the Poisson models. Ratios of added risks of AML based on both models and ratios of added risk of CML and CML plus other/unknown based on the Poisson model are much larger and cannot be estimated because the maximum likelihood estimates of the added risks are zero.

Table 8. Ratio of the lifetime added risks for an occupational BD exposure[1] concentration of 0.1 ppm based on the 95% upper confidence limit to those based on the maximum likelihood estimates for both the Poisson regression linear model and the Cox proportional hazards log linear model for three primary and three alternative definitions of subsets of leukemia (AML, CLL, and CML)

Leukemia Primary Subset	Model	Leukemia subtypes included with the primary subset			
		Primary	Primary + Other/Unknown	Primary + Unspecified	Primary + Unspecified + Other/Unknown
AML	Poisson	inf*	inf	inf	inf
	Cox	inf	inf	inf	inf
CLL	Poisson	2.2	2.1	2.2	2.1
	Cox	1.5	1.5	1.5	1.5
CML	Poisson	inf	inf	8.6	7.6
	Cox	29.9	11.4	12.1	7.7

[1]45 potential years of work with the persons being exposed between 20 and 65 years of age and ignoring exposures received 40 or more years ago and also excluding exposures received within the lag period

*inf indicates that the ratio cannot be calculated because the numerator is greater than zero and the denominator is equal to zero resulting in a value equal to infinity

Conclusions

The added risks for all leukemia mortality were calculated at an age of 85 years for an occupational exposure to BD for 45 years from ages 20 to 65 years. The calculations were made using SCOEL's methodology, European leukemia mortality rates, and European survival probabilities. The model used to calculate the added risks of all leukemia was derived by the American Chemistry Council (ACC) based on the University of Alabama (UAB) epidemiological study on workers exposed to BD. The ACC Poisson regression linear rate ratio model applied a statistical-based approach to identify variables that improved the fit of the model to the leukemia mortality observed in the UAB epidemiological data. The variables that improved significantly the likelihood of the observed leukemia mortality data were age and number of BD HITs. The final Poisson model derived by the ACC then adjusted for age and BD HITs. The final Poisson regression linear rate ratio model for leukemia with a slope of 0.000189 per occupational BD ppm-year and standard error of 0.000360 was then used to calculate added risks using SCOEL's methodology.

The same methodology used by the ACC to develop the Poisson regression model for all leukemia mortality using the UAB epidemiological data has been used here to derive both Poisson linear rate ratio models and Cox proportional hazards log-linear rate ratio models for subsets of leukemia mortality in the UAB epidemiological data. Three mutually exclusive, well-defined subsets of leukemia (AML, CLL and CML) were used as the basis to develop the models. These analyses also considered cumulative BD ppm-years as the specified dose metric in the Poisson and Cox rate ratio models. However, the cumulative BD ppm-years metric used here includes only exposures that occurred within the last 40 years of the evaluation time and excludes all exposures that were 40 years old or older as per SCOEL's request.

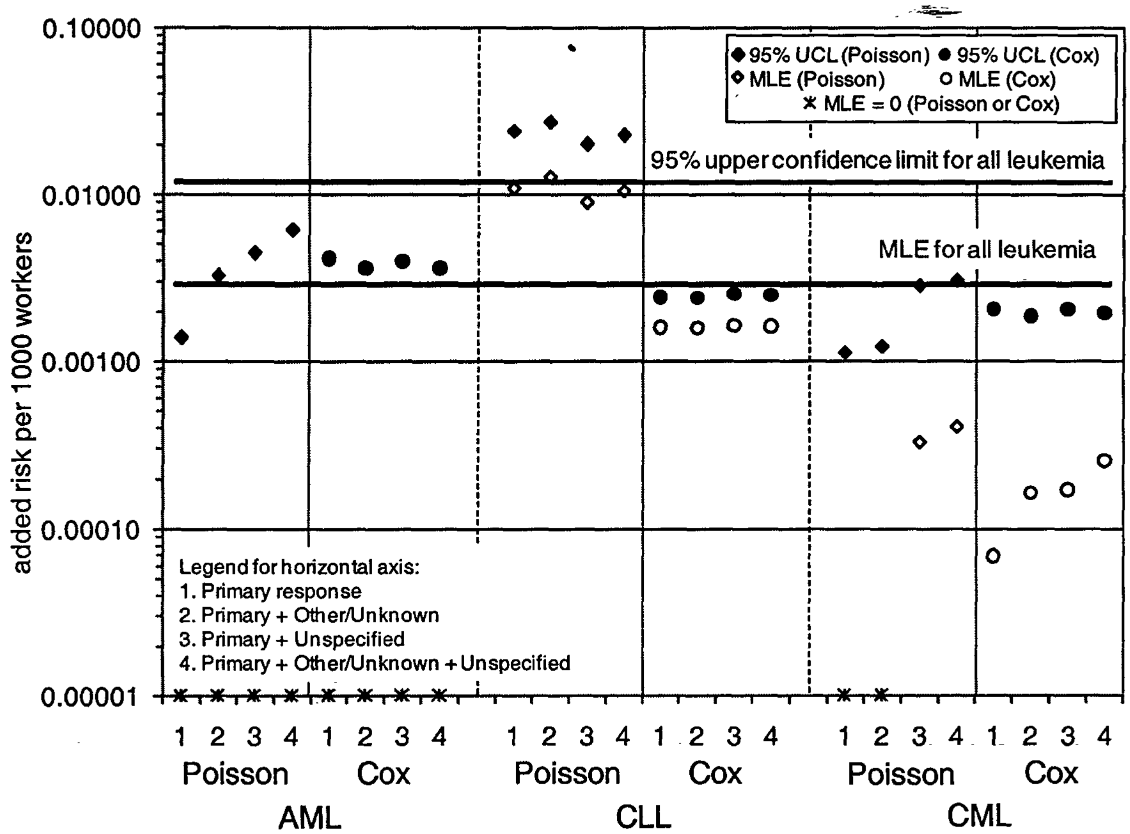
The added risks at 85 years of age for a 45-year exposure between 20 an 65 years of age to a BD concentration of 0.1 ppm at the workplace was calculated for all leukemia using the ACC maximum likelihood estimate and the corresponding 95% upper confidence limit on the slope of cumulative BD ppm-years in the Poisson regression linear model. The two horizontal lines in Figure 2 show these calculated added risks for all leukemia (0.00291 and 0.01203 per 1,000, respectively). Figure 2 also shows the added risks based on the maximum likelihood estimates and on the upper 95% confidence limits on the slopes of cumulative BD ppm-years in the Poisson and Cox models for the three endpoints (AML, CLL, and CML) and possible alternative definitions of the endpoints. All the added risks in Figure 2 are for the cumulative BD ppm-years dose metric that includes only exposures that occurred in the most recent 40 years.

In Figure 2 is clear that the added risks based on the maximum likelihood estimates for the three endpoints are smaller than the added risks for leukemia except for CLL with the Poisson regression model. These added risks for CLL with the Poisson regression model are approximately equal to the added risks for all leukemia based on the 95% upper confidence limit on the slope of the ACC model.

The added risks based on the 95% upper confidence limits on the slopes for the three endpoints are smaller than the 95% upper confidence limit of the added risk for leukemia except for CLL with the Poisson model. The upper bounds on added risks for CLL and CML and the Cox models are smaller than the MLE added risks for leukemia. This is also true for most of the CML endpoints with the Poisson model. The upper bounds on added risks for AML endpoints are by the most part between the MLE and the upper bound of the added risk of leukemia.

The leukemia mortality in the UAB epidemiological data set was fit with the Cox proportional hazards log-linear model adjusted for categorical BD HITs, similar to the Poisson model, for the sake of comparability to the analyses performed herein. The Cox regression log-linear rate ratio model for leukemia has a slope of 0.00020 per occupational BD ppm-year and standard error of 0.00013. The added risks of leukemia using this model and the same exposures specified above are 0.00308 per 1,000 for the maximum likelihood estimate and 0.00638 per 1,000 for the 95% upper confidence limit. The MLE added risk (0.00308 per 1,000) is approximately equal to the MLE added risk for leukemia using the final ACC Poisson model (0.00291 per 1,000). The upper bound of the added risk with the Cox model (0.00638 per 1,000) is approximately half the upper bound of the added risk with the final ACC model (0.01203 per 1,000).

Figure 2. Comparison of the lifetime added risks for an occupational BD exposure[1] concentration of 0.1 ppm using the maximum likelihood estimate and the 95% upper confidence limit on the slope of cumulative BD ppm-years in the Poisson regression linear model and the Cox proportional hazards log linear model for all leukemia and three primary and three alternative definitions of subsets of leukemia (AML, CLL, and CML)



[1]45 potential years of work with the persons being exposed between 20 and 65 years of age and ignoring exposures received 40 or more years ago and also excluding exposures received within the lag period

The Poisson regression linear model and the Cox proportional hazards log linear model result in comparable added risks for AML and CML, but not for CLL. The MLE added risks of CLL for the Poisson linear model are between 5.5- and 8.1-fold greater (the upper bounds on the added risks are between 7.9 and 11.4-fold greater) than for the Cox log linear model.

The discrepancies in the added risks of CLL based on the Poisson models and the Cox models are consistent with the differences in the estimated slopes of the models. All the slopes of cumulative BD ppm-years in the final Poisson regression linear rate ratio model are not different from zero at the 5% significance level. The same is true for the final Cox proportional hazards log linear model except for the model of CLL rate ratios in which the estimated slope of cumulative BD ppm-years is statistically significantly greater than zero at the

1% level. The reliance of the Poisson regression procedure on grouped data oftentimes tends to exaggerate the slope by assigning the average of the highest exposure group to much larger cumulative BD ppm-years. The Cox proportional hazards model, on the other hand, uses each individual's cumulative BD ppm-years, as opposed to grouped average exposures, and results in estimates of the dose-response relationship that are more consistent with the observed cancer mortality experience.

References

Albertini, R., E. Delzell, H. Cheng, N. Sathiakumar, M. Gargas, L. Sweeney, C. Kirman, R. Sielken, C. Valdez-Flores, J. Teta, and J. Hicks *Cancer Risk Assessment for 1,3-Butadiene (CAS No. 106-99-0)*, Prepared on behalf of the Olefins Panel, American Chemistry Council, Washington, DC.

Breslow, N.E. and N.E. Day. Statistical Methods in Cancer Research: Volume I – The Analysis of Case Control Studies, IARC Scientific Publications No. 32, Lyon France, 1980.

Checkoway, H, Pearce, N.E., and Crawford-Brown, D. J. Research Methods in Occupational Epidemiology, Oxford University Press, New York, NY, 1989.

Sielken, R.L. Jr., Valdez-Flores, C., Gargas, M.L., Kirman, C.L., Teta, M.J., and Delzell, E. (2007). Cancer risk assessment for 1,3-butadiene: dose-response modeling from an epidemiological perspective. Chemico-Biological Interactions, 166:140-149.

USEPA. (1997). Chemical and Radiation Leukemogenesis in Humans and Rodents and the Value of Rodent Models for Assessing Risks of Lymphohematopoietic Cancers. EPA/600/R-97/090

USEPA. (1998). Review of the Office of Research and Development's Draft Health Risk Assessment of 1,3-Butadiene Prepared by the Environmental Health Committee of the Science Advisory Board (SAB), EPA-SAB-EHC-99-003 (November 1998) (available at www.epa.gov/sab/pdf/ehc9903.pdf)

USEPA. (2005a). Guidelines for Carcinogen Risk Assessment. EPA/630/P-03/001B March 2005

USEPA. (2005b). Supplemental Guidance for Assessing Susceptibility from Early-Life Exposure to Carcinogens. EPA/630/R-03/003F March 2005

Appendix A

Subgroups of the Leukemia Deaths in the UAB Cohort

Table A.1. Partitioning the 81 leukemia deaths into nine subgroups

ID	BD(mm)	BD(dd)	BD(yyyy)	O(mm)	O(dd)	O(yyyy)	Race	ICD (Cause of Death)	Leukemia (Cause of Death or Contributing Cause of Death)	Subgroup
125	10	22	1903	7	22	1980	W	2059	2059	8
131	1	3	1896	12	19	1978	W	2060	2050	2
421	2	18	1943	6	18	1991	W	2080	2050	2
492	4	6	1907	9	30	1977	W	4299	2071	6
712	7	26	1904	1	26	1964	W	2041	2059	8
834	9	3	1913	1	18	1977	W	2041	2041	4
871	11	20	1897	3	9	1959	W	2042	2060	2
1026	12	26	1921	9	24	1968	W	2044	2049	7
1030	2	21	1936	1	4	1979	W	2051	2051	5
1169	8	22	1917	8	7	1967	W	2043	2050	2
1238	11	12	1925	5	11	1994	W	2050	2050	2
1474	3	14	1915	12	5	1997	W	2089	2051	5
1687	12	27	1934	6	6	1996	W	2041	2041	4
1777	5	25	1912	4	26	1974	W	2070	2040[a]	1
1778	4	16	1909	2	23	1982	W	2051	2051	5
2292	4	14	1922	4	9	1974	W	2051	2051	5
2690	7	25	1914	12	10	1970	W	2070	2040[a]	1
3010	10	31	1937	1	16	1983	W	2051	2051	5
3270	2	1	1940	9	15	1996	W	2050	2050	2
3770	3	18	1920	11	5	1998	W	2041	2041	4
4261	12	29	1923	6	25	1993	W	2059	2050	2
4907	3	16	1916	6	10	1981	W	1505	2041	4
5442	6	4	1923	10	10	1991	W	2089	2078[b]	9
5784	8	9	1918	12	17	1986	W	2059	2050	2
5811	6	12	1913	5	11	1985	B	2503	2051	5
5952	12	12	1905	12	27	1972	B	2050	2050	2
6223	8	20	1913	7	26	1993	W	2080	2050	2
6312	8	21	1912	8	30	1994	W	2050	2050	2
6347	3	9	1915	12	7	1987	W	2849	2050	2
6714	7	17	1921	2	16	1975	W	2051	2059	8
6944	5	29	1930	1	18	1976	W	2070	2070	3
7149	10	6	1929	5	22	1985	B	2041	2041	4
7166	6	19	1931	6	11	1985	W	2041	2041	4
7224	2	20	1941	3	19	1988	W	2080	2051	5
7635	9	5	1919	5	20	1998	W	2041	2041	4
7800	11	10	1955	6	1	1996	B	2050	2050	2
7987	12	22	1903	4	5	1959	W	2041	2051	5
8035	12	26	1926	6	27	1985	W	2051	2051	5
8093	1	17	1931	1	26	1989	B	2041	2041	4
8786	7	24	1940	8	29	1990	W	2041	2041	4
8791	12	19	1937	2	11	1990	W	2080	2050	2

ID	BD(mm)	BD(dd)	BD(yyyy)	O(mm)	O(dd)	O(yyyy)	Race	ICD (Cause of Death)	Leukemia (Cause of Death or Contributing Cause of Death)	Subgroup
9710	5	14	1927	10	29	1982	B	2041	2041	4
9843	4	13	1936	5	28	1995	W	2089	2070[c]	3
10057	4	2	1903	1	10	1974	B	2051	2051	5
10262	8	10	1925	3	16	1996	W	2089	2051	5
10795	8	26	1917	1	5	1976	B	2050	2050	2
10805	9	27	1929	10	29	1990	B	2080	2050	2
11111	2	26	1918	10	26	1980	W	2051	2051	5
11502	8	4	1930	1	29	1976	W	4810	2050	2
11671	11	17	1895	9	29	1972	W	2040	2040[a]	1
11711	10	19	1920	4	14	1991	W	2050	2050	2
11772	1	29	1938	6	18	1998	W	2050	2050	2
11925	12	29	1907	12	6	1971	W	2051	2051	5
11938	9	1	1912	12	26	1990	W	4149	2041	4
12297	4	10	1906	3	29	1991	W	2080	2070	3
12303	1	15	1916	5	15	1996	W	2050	2050	2
12596	10	23	1914	10	27	1984	W	2041	2041	4
12669	4	19	1891	9	10	1981	W	4140	2041	4
12932	4	11	1904	2	18	1986	W	2049	2041	4
12967	12	23	1908	6	27	1992	W	5679	2041	4
12983	6	23	1921	1	22	1967	W	2041	2050	2
13017	1	20	1926	2	14	1978	W	2050	2050	2
13199	11	28	1919	12	2	1998	W	2041	2041	4
13308	5	9	1939	1	5	1993	W	2050	2050	2
13379	1	27	1909	10	31	1971	W	2001	2041	4
13875	12	23	1924	7	20	1982	B	1629	2041	4
14103	11	21	1910	12	12	1983	W	2849	2070	3
14227	8	5	1937	12	13	1996	B	2089	2041	4
14374	6	18	1903	8	18	1980	W	2041	2041	4
14577	6	7	1940	8	5	1990	W	1629	2041	4
14909	9	17	1917	7	24	1992	W	2050	2050	2
15262	6	13	1932	3	31	1981	W	2051	2051	5
15324	10	12	1961	9	20	1994	W	2050	2050	2
15601	3	17	1916	2	5	1978	W	2041	2041	4
15604	11	10	1911	1	4	1986	W	2041	2041	4
15615	9	6	1917	11	30	1990	W	2089	2051	5
15817	2	4	1891	6	16	1972	W	2070	2050	2
15932	10	8	1933	6	11	1992	W	2049	2049	7
16078	7	22	1920	9	3	1997	W	4149	2041	4
16439	8	2	1924	8	22	1968	W	2051	2051	5
16455	9	30	1939	8	3	1991	B	2041	2041	4

[a]acute lymphocytic leukemia (ALL) excluded from analyses because there were only three cases in this cohort and also excluded from sensitivity analyses because they are not AML, CLL or CML
[b]lymphosarcoma cell leukemia excluded from sensitivity analyses because it is not AML, CLL or CML
[c]acute erythroleukemia excluded from sensitivity analyses because it is not AML, CLL or CML

Appendix B

Likelihood Functions for the Cox Proportional Hazards Models and for Grouped and Individualized Poisson Regression Models

In estimating the exposure lag for cumulative butadiene, the likelihood function of the Cox proportional hazards analyses can be used without modifications. The likelihood of the Cox proportional hazards model is based on risk sets defined for each individual with the response. The risk sets are formed according to a variable indexing the occurrence of the event (e.g., age) and, therefore, these risk sets do not change regardless of the values of the cumulative exposure or other covariate effects. Having the same risk sets, thus, guarantees that the likelihood function is calculated using the same group of individuals at the same times, even if the values for some of the variables change. Although this does not guarantee a monotone likelihood function for varying values of covariates, the fact that the risk sets are the same, offers a better behavior of the likelihood function than if the risks sets were different.

The likelihood of the Cox proportional hazards model is the sum of the likelihoods evaluated for each risk set. A risk set is defined for each worker that died with the response. The risk set is composed by all the individuals alive at the time of the response. Usually time is measured by the age of the individuals. Each individual in the risk set has its own values of the covariates at the time of the death of the worker with the response. Thus, the likelihood for the Cox proportional hazards analysis is given by:

$$\textit{Cox Likelihood} = \prod_{i=1}^{n} \left[\frac{e^{\beta x_i}}{\sum_{j=1}^{n} Y_{ij} e^{\beta x_j}} \right]^{\delta_i}$$

where n is the number of individuals in the cohort, $\delta_i = 1$ if the i-th individual died with the response and 0 otherwise, $Y_{ij} = 1$ if the age of the j-th individual is greater than or equal to the age of the i-th individual and 0 otherwise, x_i (x_j) is a vector of values of the covariates for the i-th (j-th) individual at the age the i-th individual died with the response, and β is a vector of parameters (one for each covariate) estimated using the maximum likelihood criterion.

Poisson regression analyses, on the other hand, fit models to person-years data that are split into different groups defined for each combination of the covariates included in the model. The likelihood in Poisson regression analyses uses the person years and the number of responses in each group (e.g., there is a group for each combination of covariate values or intervals) to compute the likelihood for the model. The person years and the number of responses in each group can vary with varying values of the covariates. Therefore, the likelihood for groups of person-years defined with one assumption of the covariate values does not have the same person-years and the same responses in each group of person-years defined with a different assumption of the covariate values. For example, when there is a non-zero lag for exposure, the number of person years and the number of responses in each interval of cumulative exposure are not the same as the number of person years and the number of responses in each interval of cumulative exposure when the lag is zero. The likelihood function is affected not only by the number of intervals defined for cumulative exposure and the corresponding cut points, but also by the person years and number of deaths that belong to each interval.

The likelihood for Poisson regression analyses is based on person-years that are arrayed in a table. The table usually consists of columns that represent intervals of an explanatory variable (e.g., cumulative ppm-years) and rows that are combinations of levels of covariate values (e.g., $40 < \text{age} \leq 65$, race = black, years since hire > 10). The person years in each of cells of the table satisfy each of the conditions necessary to belong to the cell – i.e., all person years in the cell (e.g., $40 < \text{age} \leq 65$, race = black, years since hire >10 and cumulative exposure between 25-100 ppm-years) correspond to black individuals when they were between 40 and 65 years of age, were first employed more than 10 years ago, and the cumulative exposure was between 25 and 100 ppm-years. If the person died from the response, then the death is counted as occurring in the cell corresponding to the person-year at the time of death. The number of deaths in each cell is then modeled as a Poisson random variable with the parameter given by the product of person years and the hazard rate per person-year in each cell. An example of a Poisson regression table is as follows:

Covariates			Cumulative exposure (ppm-years)			
			0 β_1	(0, 25] β_2	(25, 100] β_3	(100, inf) β_4
Age α_i	Race ρ_j	Years Since Hire γ_k				
≤ 40 (α_1)	Black (ρ_1)	≤ 10 (γ_1)	$n_{1,1}$ / $p_{1,1}$	$n_{1,2}$ / $p_{1,2}$	$n_{1,3}$ / $p_{1,3}$	$n_{1,4}$ / $p_{1,4}$
		>10 (γ_2)	$n_{2,1}$ / $p_{2,1}$	$n_{2,2}$ / $p_{2,2}$	$n_{2,3}$ / $p_{2,3}$	$n_{2,4}$ / $p_{2,4}$
	White (ρ_2)	≤ 10 (γ_1)	...	...	...	...
		>10 (γ_2)	...	...	...	...
	Other (ρ_3)	≤ 10 (γ_1)	...	...	...	...
		>10 (γ_2)	...	...	...	...
(40, 65] (α_2)	Black (ρ_1)	≤ 10 (γ_1)	...	...	...	...
		>10 (γ_2)	...	...	...	...
	White (ρ_2)	≤ 10 (γ_1)	...	...	...	...
		>10 (γ_2)	...	...	...	...
	Other (ρ_3)	≤ 10 (γ_1)	...	...	...	...
		>10 (γ_2)	...	...	...	...
>65 (α_3)	Black (ρ_1)	≤ 10 (γ_1)	...	...	...	...
		>10 (γ_2)	...	...	...	...
	White (ρ_2)	≤ 10 (γ_1)	...	...	...	...
		>10 (γ_2)	...	...	...	...
	Other (ρ_3)	≤ 10 (γ_1)	...	...	...	...
		>10 (γ_2)	$n_{r,1}$ / $p_{r,1}$	...	...	$n_{r,c}$ / $p_{r,c}$

The likelihood function for this Poisson table is given by:

$$\textit{Poisson Likelihood} = \prod_{i=1}^{r} \prod_{j=1}^{c} \frac{(\lambda_0 \alpha_{(i)} \rho_{(i)} \gamma_{(i)} \beta_j p_{i,j})^{n_{i,j}} \times e^{-\{\lambda_0 \alpha_{(i)} \rho_{(i)} \gamma_{(i)} \beta_j p_{i,j}\}}}{n_{i,j}!}$$

where r is the number of rows in the table, c is the number of columns, $n_{i,j}$ and $p_{i,j}$ are the number of deaths with the response and the number of person years, respectively, in the i-th row and j-th column, λ_0 is the baseline background rate of the response per person year, $\alpha_{(i)}$ is the effect of age for the i-th row of the table (here, i=1,2, ..., 18 but (i) = 1, 2 or 3) , $\rho_{(i)}$ is the effect of race for the i-th row (here, i=1,2, ..., 18 but (i) = 1, 2 or 3), $\gamma_{(i)}$ is the effect of years since hire for the i-th row (here, i=1,2, ..., 18 but (i) = 1 or 2), and β_j is the effect of the j-th interval of cumulative exposure. The parameters λ_0, $\alpha_{(i)}$, $\rho_{(i)}$, $\gamma_{(i)}$, and β_j are estimated using the maximum likelihood criterion and the values of α_1, ρ_1, γ_1, and β_1 are defined to be 1.

The likelihood function for the Poisson regression table depends on the number of person years and the number of deaths with the response in each cell.

However, the number of person years and number of deaths with the response in each cell would be different for cumulative exposures with different lags. The likelihood functional values are not comparable when the tables of person years and the number of deaths with the response are not identical. In order to alleviate this problem of non-comparable likelihood functional values, an alternative likelihood function was used when comparing the maximum likelihood procedure for alternative lag periods. The alternative likelihood function is given by:

$$\textit{Individualized Poisson Likelihood} = \prod_{i=1}^{n} \frac{(\lambda_0 \alpha_{(t)} \rho_{(t)} \gamma_{(t)} \beta_{(t)})^{\delta_i} \times e^{-\int_0^t \lambda_0 \alpha_{(s)} \rho_{(s)} \gamma_{(s)} \beta_{(s)} ds}}{\delta_i!}$$

where n is the number of workers in the cohort, t is the age of the individual when last observed, λ_0 is the baseline background rate of the response per person year, $\alpha_{(x)}$ is the effect of age at time x (i.e., although x is a continuous variable, (x) is equal to 1 if $x \leq 40$, 2 if $40 < x \leq 65$, or 3 if $x > 65$ – as indicated in the Poisson regression table), $\rho_{(x)}$ is the effect of race (i.e., (x) is equal to 1if race is black, 2 if race is white, or 3 otherwise), $\gamma_{(x)}$ is the effect of years since hire at age x (i.e., although x is a continuous variable, (x) is equal to 1 if the years since hire at age x is ≤ 10, or 2 if the years since hire at age x is >10 – as indicated in the Poisson regression table), $\beta_{(x)}$ is the effect of cumulative exposure at age x (i.e., the exposure metric can be any function of age, like cumulative exposure, and the effect of the exposure metric can be a discrete function, as that specified in the Poisson regression table, or any continuous function), and δ_i = 1if the i-th individual died with the response and 0 otherwise. This latter likelihood function is based on the individuals in the cohort. Each individual in the cohort contributes to the likelihood a fixed number of person years, regardless of the exposure lag and makes the likelihood values for different lag periods more comparable.

Appendix C

Likelihood-Based Selection of Covariates for the Poisson Regression Linear Rate Ratio Model and the Cox Proportional Hazards Log Linear Rate Ratio Model of Three Well-Defined Subsets of Leukemia (AML, CLL, and CML)

Poisson Regression Model for AML

The maximum log likelihood for AML under the null hypothesis, where the slope of cumulative BD exposure (ppm-years) in the linear model for the rate ratios is set equal to zero and there is no adjustment for confounders, is –17.68. The maximum log likelihood allowing for the possibility of a non-zero slope in the linear model for the AML rate ratio before adjusting for any of the confounders is –17.61. Using the likelihood ratio test for the hypothesis that the slope is equal to zero, the chi square statistic is 0.14 (= -2x[-17.68–(-17.61)]) with one degree of freedom corresponding to the slope being estimated in the model. This chi square value of 0.14 with one degree of freedom is not statistically significant, and indicates that a linear model with a non-zero slope for cumulative BD exposure (ppm-years) in the linear model for the AML rate ratios does not fit the observed data statistically significantly better than a linear model with zero slope. A zero slope in the linear model for the rate ratio corresponds to estimating the rate ratio as the average rate ratio instead of a linear function of cumulative BD exposure (ppm-years).

Table C.1 shows the slope in the model with the rate ratio being a linear function of cumulative BD ppm-years fit to the data after adjusting for each of the covariates. Also listed in the table are the maximum log likelihood value when the effect of the covariate is included and the maximum log likelihood value when the effect of the covariate is not included. The table also shows the chi square statistic and p-value corresponding to the hypothesis that the covariate does not improve the fit of the model. A p-value less than 0.05 implies that including the covariate in the model results in a statistically significant better fit to the data at the 5% significance level. The addition of a categorical covariate changes the partition of the person years and increases the number of terms in the likelihood function of a Poisson regression analysis. Thus, the statistically appropriate comparison is between the likelihood including the additional categorical covariate and the likelihood excluding the categorical covariate but with the same partition of person years. This can be accomplished by using the same number of terms (i.e., covariates) in the likelihood function as when the categorical covariate is included but with the value for the parameter corresponding to the added categorical covariate in the model fixed (the same value, here 1.0, for each category of the covariate – which is described in the analyses as "Covariate Excluded") as opposed to the parameter value for the covariate varying from category to category. In Table C.1, for example, the appropriate comparison is between the likelihoods for "Covariate Included" and "Covariate Excluded." For age, this is an improvement from –60.52 for "Covariate Excluded" to –47.58 for "Covariate Included" which has a chi-square statistic of $-2\times(-60.52-(-47.58))=25.88$ which has a p-value of approximately 0.0000335 in a chi-square distribution with four degrees of freedom. The four degrees of freedom (d.f.) corresponds to the number of categories (five for age minus one since the categorical covariate values are normalized.)

Table C.1. AML: Effect of each one of the non-exposure and exposure covariates on the likelihood and parameter estimates after adjusting the Poisson regression model with the rate ratio being a linear function of cumulative BD ppm-years

Covariate Considered for Inclusion in the Model	Slope of Linear Rate Ratio Model (Std Error)	Maximum Log Likelihood		Chi Square Statistic (d. f.)	p-value
		Covariate Included	Covariate Excluded		
None[1]	3.42×10^{-4} (1.01×10^{-3})	-17.61	-17.68	0.14 (1)	0.7083
Age[2]	-7.50×10^{-4} (7.09×10^{-4})	-47.58	-60.52	25.88 (4)	3.35×10^{-5}**
Years Since Hire	-1.50×10^{-4} (6.55×10^{-4})	-48.37	-55.43	14.12 (4)	0.0069**
Calendar Year	5.28×10^{-4} (1.14×10^{-3})	-48.78	-57.77	17.98 (4)	0.0012**
Race	2.03×10^{-4} (9.32×10^{-4})	-24.60	-24.82	0.44 (1)	0.5071
Plant	4.26×10^{-4} (1.08×10^{-3})	-50.83	-51.71	1.76 (5)	0.8813
STY (ppm-years)	-4.32×10^{-5} (9.19×10^{-4})	-43.97	-45.93	3.92 (5)	0.5610
DMDTC (mg/cm-years)	-3.46×10^{-4} (5.44×10^{-4})	-44.81	-49.86	10.10 (5)	0.0725
# of BD HITs	-4.40×10^{-4} (5.01×10^{-4})	-41.04	-42.74	3.40 (5)	0.6386
# of STY HITs	-5.72×10^{-4} (3.51×10^{-4})	-47.20	-52.09	9.78 (5)	0.0817
STY ≤ 50 ppm (ppm-years)	-1.23×10^{-4} (8.42×10^{-4})	-39.61	-41.85	4.48 (5)	0.4826
STY > 50 ppm (ppm-years)	-5.23×10^{-4} (4.18×10^{-4})	-44.72	-49.17	8.90 (5)	0.1131

[1]This first line compares the unadjusted model with the rate ratio being a linear function of cumulative BD ppm-years with the unadjusted null model without cumulative BD ppm-years

[2]Categories for Age, Years Since Hire and Calendar Year are based on quintiles of AML decedents, and Race is categorized as black and others, while covariates for cumulative exposures are partitioned as controls and quintiles of exposed AML decedents

*statistically significant improvement in the likelihood at the 5% significance level

**statistically significant improvement in the likelihood at the 1% significance level

The results in Table C.1 indicate that the model with the rate ratio being a linear function of cumulative BD ppm-years can fit the data statistically significantly better if the model incorporates a categorical covariate for either age, years since hire, or calendar year. Age, however, is the covariate that has the smallest p-value; i.e., adjusting for age makes a greater improvement in the fit of the model than adjusting for any of the other covariates. The slope of the AML rate ratios per ppm-year of cumulative BD exposure becomes negative when age or years since hire is included in the model but the slope increases when calendar year is included.

The best-fitting model of the AML rate ratios with the rate ratios being a linear function of cumulative BD ppm-years should include age as a covariate. The next step in the sequential consideration of covariates to include in the model has age as part of the starting model. Table C.2 lists the slope of the linear rate ratio in the model adjusted for age and one other of the non-exposure or exposure covariates. The maximum log likelihoods before and after inclusion of the one additional covariate along with the corresponding chi square statistic and p-value are also listed.

Table C.2. AML: Effect of age and each one of other non-exposure and exposure covariates on the likelihood and parameter estimates after adjusting the Poisson regression model with the rate ratio being a linear function of cumulative BD ppm-years

Covariate Considered for Inclusion in the Model	Slope of Linear Rate Ratio Model (Std Error)	Maximum Log Likelihood		Chi Square Statistic (d. f.)	p-value
		Covariate Included	Covariate Excluded		
Years Since Hire	-1.17×10^{-4} (6.79×10^{-4})	-79.41	-80.74	2.66 (4)	0.6162
Calendar Year	4.09×10^{-5} (7.98×10^{-4})	-84.33	-89.07	9.48 (4)	0.0502
Race	-1.94×10^{-4} (6.36×10^{-4})	-55.81	-56.19	0.76 (1)	0.3833
Plant	-3.38×10^{-5} (7.46×10^{-4})	-86.22	-86.91	1.38 (5)	0.9265
STY (ppm-years)	-3.22×10^{-4} (6.21×10^{-4})	-79.95	-81.69	3.48 (5)	0.6264
DMDTC (mg/cm-years)	-4.75×10^{-4} (4.35×10^{-4})	-79.25	-86.45	14.40 (5)	0.0133*
# of BD HITs	-5.81×10^{-4} (3.69×10^{-4})	-75.32	-77.55	4.46 (5)	0.4853
# of STY HITs	-6.94×10^{-4} (2.52×10^{-4})	-80.81	-86.38	11.14 (5)	0.0487*
STY ≤ 50 ppm (ppm-years)	-2.61×10^{-4} (6.97×10^{-4})	-75.34	-77.59	4.50 (5)	0.4799
STY > 50 ppm (ppm-years)	-7.12×10^{-4} (2.49×10^{-4})	-77.65	-84.07	12.84 (5)	0.0249*

*statistically significant improvement in the likelihood at the 5% significance level

Adjusting for cumulative DMDTC exposure, number of STY HITs, or cumulative STY>50 ppm in addition to adjusting for age results in a statistically significant improvement in the model fit at the 5% significance level. Cumulative DMDTC exposure, however, is the covariate that results in the largest improvement of the likelihood with a p-value of 0.0133. Adjusting for any other non-exposure covariates or exposure covariates does not result in a statistically significant improvement of the likelihood of the fit to the data. After adjusting for cumulative DMDTC, no other covariate improves significantly the fit of the log linear model to the observed AML data (Table C.3).

Table C.3. AML: Effect of age, cumulative DMDTC exposure, and each one of other non-exposure and exposure covariates on the likelihood and parameter estimates after adjusting the Poisson regression model with the rate ratio being a linear function of cumulative BD ppm-years

Covariate Considered for Inclusion in the Model	Slope of Linear Rate Ratio Model (Std Error)	Maximum Log Likelihood		Chi Square Statistic (d. f.)	p-value
		Covariate Included	Covariate Excluded		
Years Since Hire	-4.77×10^{-4} (4.32×10^{-4})	-111.59	-112.92	2.66 (4)	0.6162
Calendar Year	-4.32×10^{-4} (4.72×10^{-4})	-115.73	-119.74	8.02 (4)	0.0908
Race	-5.30×10^{-4} (3.93×10^{-4})	-89.01	-89.46	0.90 (1)	0.3428
Plant	-4.73×10^{-4} (4.42×10^{-4})	-118.39	-119.52	2.26 (5)	0.8121
STY (ppm-years)	-3.58×10^{-4} (6.04×10^{-4})	-107.95	-110.72	5.54 (5)	0.3536
# of BD HITs	-5.74×10^{-4} (3.76×10^{-4})	-106.86	-107.13	0.54 (5)	0.9906
# of STY HITs	-6.82×10^{-4} (2.69×10^{-4})	-111.92	-114.48	5.12 (5)	0.4014
STY ≤ 50 ppm (ppm-years)	-3.33×10^{-4} (6.31×10^{-4})	-104.98	-108.46	6.96 (5)	0.2236
STY > 50 ppm (ppm-years)	-7.26×10^{-4} (2.42×10^{-4})	-106.36	-109.86	7.00 (5)	0.2206

Thus, the best Poisson regression linear rate ratio model for AML should adjust for age and the cumulative exposure to DMDTC. However, as indicated at the onset of this analysis, the slope of the rate ratio is not significantly different than zero (either before or after the adjustment for these covariates) and the maximum likelihood estimate is negative for the final Poisson regression model that adjusts for statistically significant covariates.

Cox Proportional Hazards Model for AML

The Cox proportional hazards analysis is different than the Poisson regression analysis. Cox proportional hazards models intrinsically adjust for age by using age as the index variable. That is, age is not considered for inclusion because age is already part of the Cox proportional hazards model. Similarly, the slope of the cumulative BD exposure not adjusted for covariates cannot be tested against the null hypothesis of zero. However, the slope adjusted for age can be tested against the null hypothesis of zero slope. The maximum log likelihood under the null hypothesis, where the slope of cumulative BD ppm-years, adjusted for age, in the log linear model for the rate ratios is set equal to zero is –224.12. The maximum log likelihood allowing for the possibility of a non-zero slope in the linear model for the AML rate ratio before adjusting for any other of the confounders is identical up to two significant digits, i.e., –224.12. Using the likelihood ratio test for the hypothesis that the slope is equal to zero, the chi square statistic is 0.00 (= -2×[-224.12–(-224.12)]) with one degree of freedom corresponding to the slope being estimated in the model. This chi square value of 0.00 with one degree of freedom is not statistically significant, and indicates that a non-zero slope for cumulative BD ppm-years in the log linear model for the AML rate ratios does not fit the observed data statistically significantly better than a zero slope. A zero slope in the log linear model for the rate ratio corresponds to estimating the rate ratio as the average rate ratio instead of a log linear function of cumulative BD ppm-years.

Table C.4 shows the slope in the model with the rate ratio being a log linear function of cumulative BD ppm-years fit to the data before and after adjusting for each of the covariates. Also listed in the table are the maximum log likelihood value when the effect of the covariate is included and the maximum log likelihood value when the effect of the covariate is not included. The table also shows the chi square statistic and p-value corresponding to the hypothesis that the covariate does not improve the fit of the model. The addition of covariates to the Cox proportional hazards model does not change the form of the likelihood function. Thus, the maximum log likelihood value for "covariate is excluded" is identical to the maximum log likelihood value corresponding to the log linear model with no adjustment for covariates other than using age as the index variable. In Table C.4, for example, the comparison is between the likelihoods for "Covariate Included" and "Covariate Excluded." For DMDTC, this is an improvement from –224.12 for "Covariate Excluded" to –217.02 for "Covariate Included" which has a chi-square statistic value of -2×(-224.12-(-217.02))=14.20 which has a p-value of approximately 0.0144 in a chi-square distribution with five degrees of freedom. The five degrees of freedom (d.f.) corresponds to the number of categories (one for the unexposed workers and five groups for the exposed workers) minus one since the categorical covariate values are normalized.

Table C.4. AML: Effect of each one of the non-exposure and exposure covariates on the likelihood and parameter estimates after adjusting the Cox proportional hazards model with the rate ratio being a log-linear function of cumulative BD ppm-years

Covariate Considered for Inclusion in the Model	Slope of Log-Linear Rate Ratio Model (Std Error)	Maximum Log Likelihood		Chi Square Statistic (d. f.)	p-value
		Covariate Included	Covariate Excluded		
None (Age)[1]	2.54×10^{-5} (3.89×10^{-4})	-224.12	-224.12	0.00 (4)	1.0000
Years Since Hire[2]	2.95×10^{-5} (3.97×10^{-4})	-223.24	-224.12	1.77 (4)	0.7783
Calendar Year	4.45×10^{-5} (3.67×10^{-4})	-221.78	-224.12	4.68 (4)	0.3213
Race	-6.79×10^{-5} (4.18×10^{-4})	-223.76	-224.12	0.72 (1)	0.3971
Plant	1.66×10^{-4} (3.73×10^{-4})	-219.18	-224.12	9.88 (5)	0.0786
STY (ppm-years)	1.85×10^{-5} (4.75×10^{-4})	-222.53	-224.12	3.19 (5)	0.6712
DMDTC (mg/cm-years)	-2.20×10^{-4} (5.67×10^{-4})	-217.02	-224.12	14.20 (5)	0.0144*
# of BD HITs	-3.07×10^{-4} (7.13×10^{-4})	-222.05	-224.12	4.15 (5)	0.5283
# of STY HITs	-6.65×10^{-4} (7.74×10^{-4})	-218.41	-224.12	11.43 (5)	0.0435*
STY ≤ 50 ppm (ppm-years)	2.46×10^{-5} (4.55×10^{-4})	-221.85	-224.12	4.54 (5)	0.4748
STY > 50 ppm (ppm-years)	-6.03×10^{-4} (7.79×10^{-4})	-218.51	-224.12	11.22 (5)	0.0472*

[1]This first line compares the model with the rate ratio being a log linear function of cumulative BD ppm-years (adjusted only for age) with the null model without cumulative BD ppm-years

[2]Categories for Years Since Hire and Calendar Year are based on quintiles of AML decedents, and Race is categorized as black and others, while covariates for cumulative exposures are partitioned as controls and quintiles of exposed AML decedents

*statistically significant improvement in the likelihood at the 5% significance level

The results in Table C.4 indicate that the model with the rate ratio being a log linear function of cumulative BD ppm-years can fit the data statistically significantly better if the model incorporates a categorical covariate for either cumulative DMDTC, number of STY HITs, or cumulative ppm-years of STY>50 ppm. Cumulative DMDTC exposure, however, is the covariate that has the smallest p-value; i.e., adjusting for cumulative DMDTC makes a greater improvement in the fit of the model than adjusting for any of the other covariates.

The slope of the AML rate ratios per ppm-year of cumulative BD exposure becomes negative when any of the statistically significant covariates is included in the model. After adjusting for cumulative DMDTC, no other covariate improves significantly the fit of the log linear model to the observed AML data (Table C.5).

Table C.5. AML: Effect of cumulative DMDTC exposure and each one of other non-exposure and exposure covariates on the likelihood and parameter estimates after adjusting the Cox proportional hazards model with the rate ratio being a log-linear function of cumulative BD ppm-years

Covariate Considered for Inclusion in the Model	Slope of Log-Linear Rate Ratio Model (Std Error)	Maximum Log Likelihood		Chi Square Statistic (d. f.)	p-value
		Covariate Included	Covariate Excluded		
Years Since Hire[12]	-2.05×10^{-4} (5.69×10^{-4})	-215.95	-217.02	2.14 (4)	0.7102
Calendar Year	-1.94×10^{-4} (5.40×10^{-4})	-215.54	-217.02	2.96 (4)	0.5639
Race	-3.36×10^{-4} (5.93×10^{-4})	-216.58	-217.02	0.89 (1)	0.3452
Plant	-1.56×10^{-5} (5.15×10^{-4})	-212.91	-217.02	8.24 (5)	0.1436
STY (ppm-years)	3.00×10^{-7} (5.06×10^{-4})	-214.02	-217.02	6.00 (5)	0.3064
# of BD HITs	-3.23×10^{-4} (7.12×10^{-4})	-216.57	-217.02	0.90 (5)	0.9701
# of STY HITs	-6.21×10^{-4} (8.00×10^{-4})	-214.18	-217.02	5.69 (5)	0.3376
STY ≤ 50 ppm (ppm-years)	-1.57×10^{-5} (4.96×10^{-4})	-213.25	-217.02	7.55 (5)	0.1831
STY > 50 ppm (ppm-years)	-6.69×10^{-4} (8.46×10^{-4})	-214.05	-217.02	5.95 (5)	0.3109

Thus, the best Cox proportional hazards log linear rate ratio model for AML should adjust for the cumulative exposure to DMDTC in addition to using age as the index variable. However, as indicated at the onset of this analysis, the slope of the log linear rate ratio is not significantly different than zero (either before or after adjustment for these covariates) and the maximum likelihood estimate is negative for the final Cox proportional hazards model that adjusts for statistically significant covariates.

Poisson Regression Model for CLL

The maximum log likelihood for CLL under the null hypothesis, where the slope of cumulative BD ppm-years in the linear model for the rate ratios is set equal to zero and there is no adjustment for confounders, is –22.70. The maximum log likelihood allowing for the possibility of a non-zero slope in the linear model for the CLL rate ratio before adjusting for any of the confounders is –18.34. Using the likelihood ratio test for the hypothesis that the slope is equal to zero, the chi square statistic is 8.72 (=-2x[-22.70–(-18.34)]) with one degree of freedom. This chi square value of 8.72 is statistically significant at the 1% level, and indicates that a non-zero slope for cumulative BD ppm-years in the linear model for the CLL rate ratios fit the observed data statistically significantly better than a zero slope (when there are no covariates in the model).

Table C.6 shows the slope in the model with the rate ratio being a linear function of cumulative BD ppm-years fit to the data after adjusting for each of the covariates. The results indicate that the model with the rate ratio being a linear function of cumulative BD ppm-years can fit the data statistically significantly better if the model incorporates a categorical covariate for either age, years since hire, calendar year, or plant. Age, however, is the covariate that has the smallest p-value. The slope of the CLL rate ratios per ppm-year of cumulative BD exposure decrease by 50% and 80% when age or years since hire, respectively, is included in the model but the slope increases by approximately 6% and 50% when calendar year or plant, respectively, is added to the model.

Table C.6. CLL: Effect of each one of the non-exposure and exposure covariates on the likelihood and parameter estimates after adjusting the Poisson regression model with the rate ratio being a linear function of cumulative BD ppm-years

Covariate Considered for Inclusion in the Model	Slope of Linear Rate Ratio Model (Std Error)	Maximum Log Likelihood		Chi Square Statistic (d. f.)	p-value
		Covariate Included	Covariate Excluded		
None[1]	4.37×10^{-3} (2.85×10^{-3})	-18.34	-22.70	8.72 (1)	0.0031**
Age[2]	2.85×10^{-3} (2.08×10^{-3})	-47.22	-66.08	37.72 (4)	1.28×10^{-7}**
Years Since Hire	2.42×10^{-3} (1.85×10^{-3})	-51.75	-65.01	26.52 (4)	2.49×10^{-5}**
Calendar Year	4.60×10^{-3} (2.98×10^{-3})	-50.10	-58.38	16.56 (4)	0.0024**
Race	3.41×10^{-3} (2.55×10^{-3})	-26.87	-27.34	0.94 (1)	0.3323
Plant	6.43×10^{-3} (4.04×10^{-3})	-48.75	-54.43	11.36 (5)	0.0447*
STY (ppm-years)	4.06×10^{-3} (4.36×10^{-3})	-39.09	-39.74	1.30 (5)	0.9349
DMDTC (mg/cm-years)	2.36×10^{-3} (2.16×10^{-3})	-46.81	-48.82	4.02 (5)	0.5465
# of BD HITs	2.52×10^{-3} (2.66×10^{-3})	-41.34	-43.89	5.10 (5)	0.4038
# of STY HITs	4.90×10^{-3} (4.57×10^{-3})	-43.21	-43.63	0.84 (5)	0.9744
STY ≤ 50 ppm (ppm-years)	1.81×10^{-3} (2.27×10^{-3})	-38.15	-39.32	2.34 (5)	0.8004
STY > 50 ppm (ppm-years)	4.13×10^{-3} (4.14×10^{-3})	-43.01	-46.68	7.34 (5)	0.1966

[1]This first line compares the unadjusted model with the rate ratio being a linear function of cumulative BD ppm-years with the unadjusted null model without cumulative BD ppm-years
[2]Categories for Age, Years Since Hire and Calendar Year were based on quintiles of CLL decedents, and Race was categorized as black and others, while covariates for cumulative exposures were partitioned as controls and quintiles of exposed CLL decedents
*statistically significant improvement in the likelihood at the 5% significance level
**statistically significant improvement in the likelihood at the 1% significance level

The best-fitting model of the CLL rate ratios with the rate ratios being a linear function of cumulative BD ppm-years should include age as a covariate. The next step in the sequential consideration of covariates to include in the model has age as part of the starting model. Adjusting for any other non-exposure covariates or exposure covariates does not result in a statistically significant improvement of the likelihood of the fit to the data. After adjusting for age, no other covariate improves significantly the fit of the linear model to the observed CLL data (Table C.7).

Table C.7. CLL: Effect of age and each one of other non-exposure and exposure covariates on the likelihood and parameter estimates after adjusting the Poisson regression model with the rate ratio being a linear function of cumulative BD ppm-years

Covariate Considered for Inclusion in the Model	Slope of Linear Rate Ratio Model (Std Error)	Maximum Log Likelihood		Chi Square Statistic (d. f.)	p-value
		Covariate Included	Covariate Excluded		
Years Since Hire	2.64×10^{-3} (1.98×10^{-3})	-82.22	-83.62	2.80 (4)	0.5918
Calendar Year	2.97×10^{-3} (2.16×10^{-3})	-86.21	-89.52	6.62 (4)	0.1574
Race	1.88×10^{-3} (1.72×10^{-3})	-59.64	-60.43	1.58 (1)	0.2088
Plant	3.74×10^{-3} (2.61×10^{-3})	-83.33	-88.00	9.34 (5)	0.0962
STY (ppm-years)	2.29×10^{-3} (2.73×10^{-3})	-73.32	-74.61	2.58 (5)	0.7644
DMDTC (mg/cm-years)	1.79×10^{-3} (1.77×10^{-3})	-81.74	-83.53	3.58 (5)	0.6113
# of BD HITs	1.50×10^{-3} (1.77×10^{-3})	-75.81	-78.53	5.44 (5)	0.3646
# of STY HITs	2.23×10^{-3} (2.42×10^{-3})	-76.63	-77.32	1.38 (5)	0.9265
STY ≤ 50 ppm (ppm-years)	1.52×10^{-3} (1.92×10^{-3})	-71.77	-73.37	3.20 (5)	0.6692
STY > 50 ppm (ppm-years)	1.10×10^{-3} (1.66×10^{-3})	-77.36	-79.96	5.20 (5)	0.3920

Thus, the best Poisson regression linear rate ratio model for CLL should adjust for age. However, even though the slope of the rate ratio is significantly different than zero before any adjustment for covariates, the maximum likelihood estimate of the slope for cumulative BD ppm-years in the final Poisson regression model that adjusts for age is not statistically significantly different from zero using Wald's test.

Cox Proportional Hazards Model for CLL

The maximum log likelihood for CLL under the null hypothesis, where the slope of cumulative BD ppm-years, adjusted for age, in the log linear model for the rate ratios is set equal to zero is –208.39. The maximum log likelihood allowing for the possibility of a non-zero slope in the log linear model for the CLL rate ratio before adjusting for any other of the confounders is –205.83. Using the likelihood ratio test for the hypothesis that the slope is equal to zero, the chi square statistic is 5.12 (=-2×[-208.39–(-205.83)]) with one degree of freedom corresponding to the slope being estimated in the model. This chi square value of 5.12 with one degree of freedom is statistically significant at the 5% significance level, and indicates that a non-zero slope for cumulative BD ppm-years in the log linear model for the CLL rate ratios fit the observed data statistically significantly better than a zero slope (when there are no covariates in the model).

Table C.8 shows the slope in the model with the rate ratio being a log linear function of cumulative BD ppm-years fit to the data before and after adjusting for each of the covariates.

Table C.8. CLL: Effect of each one of the non-exposure and exposure covariates on the likelihood and parameter estimates after adjusting the Cox proportional hazards model with the rate ratio being a log-linear function of cumulative BD ppm-years

Covariate Considered for Inclusion in the Model	Slope of Log-Linear Rate Ratio Model (Std Error)	Maximum Log Likelihood		Chi Square Statistic (d. f.)	p-value
		Covariate Included	Covariate Excluded		
None (Age)[1]	4.15×10^{-4} (1.32×10^{-4})	-205.83	-208.39	5.12 (1)	0.0237*
Years Since Hire[2]	4.02×10^{-4} (1.33×10^{-4})	-205.19	-205.83	1.29 (4)	0.8626
Calendar Year	3.94×10^{-4} (1.30×10^{-4})	-202.46	-205.83	6.76 (4)	0.1494
Race	3.11×10^{-4} (1.54×10^{-4})	-204.95	-205.83	1.77 (1)	0.1829
Plant	5.09×10^{-4} (1.61×10^{-4})	-200.99	-205.83	9.68 (5)	0.0848
STY (ppm-years)	4.08×10^{-4} (1.73×10^{-4})	-203.77	-205.83	4.14 (5)	0.5299
DMDTC (mg/cm-years)	4.00×10^{-4} (1.58×10^{-4})	-202.88	-205.83	5.90 (5)	0.3160
# of BD HITs	3.66×10^{-4} (1.53×10^{-4})	-201.88	-205.83	7.91 (5)	0.1613
# of STY HITs	3.56×10^{-4} (1.45×10^{-4})	-203.80	-205.83	4.07 (5)	0.5391
STY ≤ 50 ppm (ppm-years)	4.07×10^{-4} (1.73×10^{-4})	-203.32	-205.83	5.03 (5)	0.4125
STY > 50 ppm (ppm-years)	3.68×10^{-4} (1.92×10^{-4})	-202.23	-205.83	7.21 (5)	0.2053

[1]This first line compares the model with the rate ratio being a log linear function of cumulative BD ppm-years (adjusted only for age) with the null model without cumulative BD ppm-years

[2]Categories for Years Since Hire and Calendar Year were based on quintiles of CLL decedents, and Race was categorized as black and others, while covariates for cumulative exposures were partitioned as controls and quintiles of exposed CLL decedents

*statistically significant improvement in the likelihood at the 5% significance level

The results in Table C.8 indicate that the fit of the model with the rate ratio being a log linear function of cumulative BD ppm-years does not improve significantly if the model adjusts for any of the ten covariates considered. Thus, the best Cox proportional hazards log linear rate ratio model for CLL, which implicitly adjusts for age using it as the index variable, does not need to adjust for any of the ten covariates considered. The slope of the final log linear model for CLL is significantly greater than zero at the 5% significance level using Wald's test.

Poisson Regression Model for CML

The maximum log likelihood for CML under the null hypothesis, where the slope of cumulative BD ppm-years in the linear model for the rate ratios is set equal to zero and there is no adjustment for confounders, is –18.44. The maximum log likelihood allowing for the possibility of a non-zero slope in the linear model for the CML rate ratio before adjusting for any of the confounders is –16.18. Using the likelihood ratio test for the hypothesis that the slope is equal to zero, the chi square statistic is 4.52 (=-2×[-18.44–(-16.18)]) with one degree of freedom. This chi square value of 4.52 is statistically significant at the 5% level, and indicates that a linear model with a non-zero slope for cumulative BD ppm-years in the linear model for the CML rate ratios fit the observed data statistically significantly better than a zero slope (when there are no covariates in the model).

Table C.9 shows the slope in the model with the rate ratio being a linear function of cumulative BD ppm-years fit to the data after adjusting for each of the covariates. The results indicate that the model with the rate ratio being a linear function of cumulative BD ppm-years can fit the data statistically significantly better if the model incorporates a categorical covariate for either years since hire, or the number of BD HITs. Years since hire, however, is the covariate that has the smallest p-value. The slope of the CML rate ratios per ppm-year of cumulative BD exposure decreases by a factor of about two when years since hire is included in the model and decreases by a factor of 15 when the number of BD HITs is added to the model.

Table C.9. CML: Effect of each one of the non-exposure and exposure covariates on the likelihood and parameter estimates after adjusting the Poisson regression model with the rate ratio being a linear function of cumulative BD ppm-years

Covariate Considered for Inclusion in the Model	Slope of Linear Rate Ratio Model (Std Error)	Maximum Log Likelihood		Chi Square Statistic (d. f.)	p-value
		Covariate Included	Covariate Excluded		
None[1]	3.62×10^{-3} (3.21×10^{-3})	-16.18	-18.44	4.52 (1)	0.0335*
Age[2]	2.43×10^{-3} (2.38×10^{-3})	-38.98	-43.50	9.04 (4)	0.0601
Years Since Hire	1.84×10^{-3} (1.94×10^{-3})	-39.42	-45.25	11.66 (4)	0.0201*
Calendar Year	3.43×10^{-3} (3.09×10^{-3})	-38.71	-39.69	1.96 (4)	0.7431
Race	4.23×10^{-3} (3.88×10^{-3})	-21.82	-21.94	0.24 (1)	0.6242
Plant	3.77×10^{-3} (3.46×10^{-3})	-41.45	-42.25	1.60 (5)	0.9012
STY (ppm-years)	1.60×10^{-3} (2.48×10^{-3})	-31.94	-34.31	4.74 (5)	0.4484
DMDTC (mg/cm-years)	1.07×10^{-3} (1.49×10^{-3})	-37.19	-41.40	8.42 (5)	0.1346
# of BD HITs	2.37×10^{-4} (7.73×10^{-4})	-32.37	-38.39	12.04 (5)	0.0342*
# of STY HITs	6.52×10^{-4} (1.13×10^{-3})	-35.58	-38.32	5.48 (5)	0.3601
STY ≤ 50 ppm (ppm-years)	5.93×10^{-4} (1.28×10^{-3})	-31.18	-33.20	4.04 (5)	0.5437
STY > 50 ppm (ppm-years)	3.11×10^{-3} (3.43×10^{-3})	-38.43	-41.08	5.30 (5)	0.3804

[1]This first line compares the unadjusted model with the rate ratio being a linear function of cumulative BD ppm-years with the unadjusted null model without cumulative BD ppm-years

[2]Categories for Age, Years Since Hire and Calendar Year were based on quintiles of CML decedents, and Race was categorized as black and others, while covariates for cumulative exposures were partitioned as controls and quintiles of exposed CML decedents

*statistically significant improvement in the likelihood at the 5% significance level

The best-fitting model of the CML rate ratios with the rate ratios being a linear function of cumulative BD ppm-years should include years since hire as a covariate. The next step in the sequential consideration of covariates is to adjust the model for years since hire and use it as the starting model. Table C.10 lists the slope of the linear rate ratio in the model adjusted for years since hire and one other of the non-exposure or exposure covariates. The maximum log likelihoods after inclusion of the one additional covariate along with the corresponding chi square statistic and p-value are also listed.

Table C.10. CML: Effect of years since hire and each one of other non-exposure and exposure covariates on the likelihood and parameter estimates after adjusting the Poisson regression model with the rate ratio being a linear function of cumulative BD ppm-years

Covariate Considered for Inclusion in the Model	Slope of Linear Rate Ratio Model (Std Error)	Maximum Log Likelihood		Chi Square Statistic (d. f.)	p-value
		Covariate Included	Covariate Excluded		
Age	1.93×10^{-3} (2.01×10^{-3})	-60.74	-62.54	3.60 (4)	0.4628
Calendar Year	1.54×10^{-3} (1.73×10^{-3})	-57.63	-58.85	2.44 (4)	0.6554
Race	1.96×10^{-3} (2.17×10^{-3})	-47.35	-47.37	0.04 (1)	0.8415
Plant	1.76×10^{-3} (1.94×10^{-3})	-68.10	-68.87	1.54 (5)	0.9084
STY (ppm-years)	9.94×10^{-4} (1.71×10^{-3})	-57.67	-59.72	4.10 (5)	0.5351
DMDTC (mg/cm-years)	6.80×10^{-4} (1.14×10^{-3})	-61.97	-65.55	7.16 (5)	0.2090
# of BD HITs	1.11×10^{-4} (6.28×10^{-4})	-58.49	-64.16	11.34 (5)	0.0450*
# of STY HITs	2.90×10^{-4} (7.73×10^{-4})	-63.69	-66.23	5.08 (5)	0.4062
STY ≤ 50 ppm (ppm-years)	4.95×10^{-4} (1.14×10^{-3})	-57.14	-58.73	3.18 (5)	0.6723
STY > 50 ppm (ppm-years)	1.39×10^{-3} (1.87×10^{-3})	-64.40	-66.74	4.68 (5)	0.4562

*statistically significant improvement in the likelihood at the 5% significance level

Table C.10 shows that adjusting for the number of BD HITs in addition to adjusting for years since hire results in a statistically significant improvement in the model fit at the 5% significance level. Adjusting for any other non-exposure covariates or exposure covariates does not result in a statistically significant improvement of the likelihood of the fit to the data. After adjusting for years since hire and the number of BD HITs no other covariate improves significantly the fit of the log linear model to the observed CML data (Table C.11).

Table C.11. CML: Effect of years since hire, number of BD HITs, and each one of other non-exposure and exposure covariates on the likelihood and parameter estimates after adjusting the Poisson regression model with the rate ratio being a linear function of cumulative BD ppm-years

Covariate Considered for Inclusion in the Model	Slope of Linear Rate Ratio Model (Std Error)	Maximum Log Likelihood		Chi Square Statistic (d. f.)	p-value
		Covariate Included	Covariate Excluded		
Age	8.54×10^{-5} (5.93×10^{-4})	-78.96	-81.19	4.46 (4)	0.3473
Calendar Year	6.28×10^{-5} (5.76×10^{-4})	-76.74	-77.94	2.40 (4)	0.6626
Race	3.43×10^{-5} (5.53×10^{-4})	-63.38	-63.48	0.20 (1)	0.6547
Plant	1.02×10^{-4} (6.26×10^{-4})	-84.58	-85.23	1.30 (5)	0.9349
STY (ppm-years)	1.55×10^{-4} (7.35×10^{-4})	-77.07	-78.59	3.04 (5)	0.6938
DMDTC (mg/cm-years)	8.61×10^{-5} (5.91×10^{-4})	-83.85	-86.06	4.42 (5)	0.4907
# of STY HITs	-4.21×10^{-5} (4.62×10^{-4})	-85.12	-85.85	1.46 (5)	0.9176
STY ≤ 50 ppm (ppm-years)	-1.87×10^{-5} (5.00×10^{-4})	-77.64	-78.46	1.64 (5)	0.8964
STY > 50 ppm (ppm-years)	2.59×10^{-4} (8.98×10^{-4})	-83.51	-84.25	1.48 (5)	0.9154

Thus, the best Poisson regression linear rate ratio model for CML should adjust for years since hire and the number of BD HITs. However, although the slope of the rate ratio is significantly greater than zero before any adjustment for covariates the maximum likelihood estimate is not statistically significantly different from zero for the final Poisson regression model that adjusts for statistically significant covariates.

Although the best Poisson regression linear rate ratio model for CML adjusts for years since hire first (as suggested in Table C.9) and the number of BD HITs second, an alternative model would be to choose age first in Table C.9 (age is close to being significant at the 5% level with a p-value of 0.0601). There are at least three reasons why the epidemiological model for CML mortality should be adjusted for age a priori: 1) age is known to be associated with increased CML mortality, 2) the statistical significance of years since hire is at least partially related to the age of the workers, and 3) the results of the Poisson regression model adjusted for age are more comparable to the Cox proportional hazards model that adjusts for age by using it as the index variable. If age is selected as the first covariate to adjust the CML rate ratio model, the next step in the sequential consideration of covariates to include in the model has age as part of the starting model. Table C.12 lists the slope of the linear rate ratio in the model adjusted for age and one other of the non-exposure or exposure covariates. The maximum log likelihoods after inclusion of the one additional covariate along with the corresponding chi square statistic and p-value are also listed

Table C.12. CML: Effect of age and each one of other non-exposure and exposure covariates on the likelihood and parameter estimates after adjusting the Poisson regression model with the rate ratio being a linear function of cumulative BD ppm-years

Covariate Considered for Inclusion in the Model	Slope of Linear Rate Ratio Model (Std Error)	Maximum Log Likelihood		Chi Square Statistic (d. f.)	p-value
		Covariate Included	Covariate Excluded		
Years Since Hire	1.93×10^{-3} (2.01×10^{-3})	-60.74	-63.85	6.22 (4)	0.1833
Calendar Year	2.23×10^{-3} (2.25×10^{-3})	-63.78	-64.54	1.52 (4)	0.8231
Race	2.79×10^{-3} (2.84×10^{-3})	-46.74	-46.80	0.12 (1)	0.7290
Plant	2.42×10^{-3} (2.45×10^{-3})	-65.97	-66.71	1.48 (5)	0.9154
STY (ppm-years)	1.04×10^{-3} (1.77×10^{-3})	-56.81	-59.35	5.08 (5)	0.4062
DMDTC (mg/cm-years)	8.74×10^{-4} (1.33×10^{-3})	-61.11	-65.39	8.56 (5)	0.1280
# of BD HITs	1.28×10^{-4} (6.39×10^{-4})	-57.38	-63.74	12.72 (5)	0.0261*
# of STY HITs	3.10×10^{-4} (8.02×10^{-4})	-61.02	-64.20	6.36 (5)	0.2727
STY ≤ 50 ppm (ppm-years)	5.19×10^{-4} (1.17×10^{-3})	-55.90	-57.87	3.94 (5)	0.5581
STY > 50 ppm (ppm-years)	1.60×10^{-3} (2.10×10^{-3})	-62.68	-65.59	5.82 (5)	0.3241

*statistically significant improvement in the likelihood at the 5% significance level

Similarly to when years since hire was selected as the first covariate in the model, Table C.12 shows that adjusting for the number of BD HITs in addition to adjusting for age results in a statistically significant improvement in the model fit at the 5% significance level. Adjusting for any other non-exposure covariates or exposure covariates does not result in a statistically significant improvement of the likelihood of the fit to the data. After adjusting for age and the number of BD HITs, no other covariate improves significantly the fit of the log linear model to the observed CML data (Table C.13).

Table C.13. CML: Effect of age, number of BD HITs, and each one of other non-exposure and exposure covariates on the likelihood and parameter estimates after adjusting the Poisson regression model with the rate ratio being a linear function of cumulative BD ppm-years

Covariate Considered for Inclusion in the Model	Slope of Linear Rate Ratio Model (Std Error)	Maximum Log Likelihood		Chi Square Statistic (d. f.)	p-value
		Covariate Included	Covariate Excluded		
Years Since Hire	8.54×10^{-5} (5.93×10^{-4})	-78.96	-81.81	5.70 (4)	0.2227
Calendar Year	8.58×10^{-5} (5.96×10^{-4})	-82.52	-83.25	1.46 (4)	0.8337
Race	5.74×10^{-5} (5.76×10^{-4})	-61.76	-61.83	0.14 (1)	0.7083
Plant	1.24×10^{-4} (6.43×10^{-4})	-82.30	-82.88	1.16 (5)	0.9487
STY (ppm-years)	1.30×10^{-4} (6.91×10^{-4})	-74.86	-76.59	3.46 (5)	0.6294
DMDTC (mg/cm-years)	1.15×10^{-4} (6.17×10^{-4})	-81.68	-84.16	4.96 (5)	0.4208
# of STY HITs	-3.02×10^{-5} (4.69×10^{-4})	-81.71	-82.55	1.68 (5)	0.8914
STY ≤ 50 ppm (ppm-years)	-3.81×10^{-5} (4.72×10^{-4})	-75.99	-76.87	1.76 (5)	0.8813
STY > 50 ppm (ppm-years)	2.47×10^{-4} (8.70×10^{-4})	-80.35	-81.20	1.70 (5)	0.8889

Thus, the best Poisson regression linear rate ratio model with age as a covariate for CML should also adjust for the number of BD HITs. However, although the slope of the rate ratio is significantly greater than zero before any adjustment for covariates, the maximum likelihood estimate of the slope for cumulative BD ppm-years is not statistically significantly different from zero for the final Poisson regression model that adjusts for age and other statistically significant covariates.

The estimated slope (MLE=1.11×10^{-4}, S.E.=6.28×10^{-4}) of the Poisson regression model for CML that adjusts for the most significant covariates (years since hire and number of BD HITs) is not statistically significantly different than zero and is slightly smaller than the slope estimated (MLE=1.28×10^{-4} S.E.=6.39×10^{-4}) for the model that adjusts for age and other significant covariates (number of BD HITs) which is also not statistically significantly different than zero.

Cox Proportional Hazards Model for CML

The maximum log likelihood for CML under the null hypothesis, where the slope of cumulative BD ppm-years, adjusted for age, in the log linear model for the rate ratios is set equal to zero is –141.14. The maximum log likelihood allowing for the possibility of a non-zero slope in the log-linear model for the CML rate ratio before adjusting for any other of the confounders is –140.94. Using the likelihood ratio test for the hypothesis that the slope is equal to zero, the chi square statistic is 0.40 (=-2×[-141.14–(-140.94)]) with one degree of freedom corresponding to the slope being estimated in the model. This chi square value of 0.40 with one degree of freedom is not statistically significant, and indicates that a non-zero slope for cumulative BD ppm-years in the log linear model for the CML rate ratios does not fit the observed data statistically significantly better than a zero slope. A zero slope in the log linear model for the rate ratio corresponds to estimating the rate ratio as the average rate ratio instead of a log linear function of cumulative BD ppm-years.

Table C.14 shows the slope in the model with the rate ratio being a log linear function of cumulative BD ppm-years fit to the data after adjusting for each of the covariates.

Table C.14. CML: Effect of each one of the non-exposure and exposure covariates on the likelihood and parameter estimates after adjusting the Cox proportional hazards model with the rate ratio being a log-linear function of cumulative BD ppm-years

Covariate Considered for Inclusion in the Model	Slope of Log-Linear Rate Ratio Model (Std Error)	Maximum Log Likelihood		Chi Square Statistic (d. f.)	p-value
		Covariate Included	Covariate Excluded		
None (Age)[1]	2.20E-4 (2.88E-4)	-140.94	-141.14	0.40 (1)	0.5271
Years Since Hire[2]	2.00×10^{-4} (3.01×10^{-4})	-138.39	-140.94	5.12 (4)	0.2757
Calendar Year	2.08×10^{-4} (3.00×10^{-4})	-138.73	-140.94	4.42 (4)	0.3521
Race	2.11×10^{-4} (3.15×10^{-4})	-140.94	-140.94	0.01 (1)	0.9436
Plant	2.87×10^{-4} (3.41×10^{-4})	-139.16	-140.94	3.56 (5)	0.6143
STY (ppm-years)	6.47×10^{-5} (4.23×10^{-4})	-137.71	-140.94	6.47 (5)	0.2635
DMDTC (mg/cm-years)	6.87×10^{-5} (3.79×10^{-4})	-135.97	-140.94	9.94 (5)	0.0769
# of BD HITs	-1.29×10^{-4} (6.08×10^{-4})	-133.46	-140.94	14.98 (5)	0.0105*
# of STY HITs	-7.92×10^{-5} (4.67×10^{-4})	-136.68	-140.94	8.52 (5)	0.1298
STY ≤ 50 ppm (ppm-years)	-3.12×10^{-5} (4.64×10^{-4})	-138.11	-140.94	5.67 (5)	0.3396
STY > 50 ppm (ppm-years)	1.53×10^{-4} (3.77×10^{-4})	-137.44	-140.94	7.00 (5)	0.2209

[1]This first line compares the model with the rate ratio being a log linear function of cumulative BD ppm-years (adjusted only for age) with the null model without cumulative BD ppm-years

[2]Categories for Years Since Hire and Calendar Year were based on quintiles of CML decedents, and Race was categorized as black and others, while covariates for cumulative exposures were partitioned as controls and quintiles of exposed CML decedents

*statistically significant improvement in the likelihood at the 5% significance level

The results in Table C.14 indicate that the model with the rate ratio being a log linear function of cumulative BD ppm-years can fit the data statistically significantly better if the model incorporates a categorical covariate for the number of BD HITs. The slope of the CML rate ratios per ppm-year of cumulative BD exposure becomes negative when the number of BD HITs is included in the model. After adjusting for the number of BD HITs, no other covariate improves significantly the fit of the log linear model to the observed CML data (Table C.15).

Table C.15. CML: Effect of number of BD HITs and each one of other non-exposure and exposure covariates on the likelihood and parameter estimates after adjusting the Cox proportional hazards model with the rate ratio being a log-linear function of cumulative BD ppm-years

Covariate Considered for Inclusion in the Model	Slope of Log-Linear Rate Ratio Model (Std Error)	Maximum Log Likelihood		Chi Square Statistic (d. f.)	p-value
		Covariate Included	Covariate Excluded		
Years Since Hire	-1.29×10^{-4} (6.08×10^{-4})	-131.36	-133.46	4.19 (4)	0.3804
Calendar Year	-1.57×10^{-4} (6.28×10^{-4})	-131.50	-133.46	3.91 (4)	0.4185
Race	-1.69×10^{-4} (6.49×10^{-4})	-133.34	-133.46	0.24 (1)	0.6242
Plant	-2.18×10^{-4} (6.57×10^{-4})	-132.06	-133.46	2.79 (5)	0.7317
STY (ppm-years)	-1.76×10^{-4} (6.86×10^{-4})	-131.79	-133.46	3.32 (5)	0.6506
DMDTC (mg/cm-years)	-1.84×10^{-4} (7.48×10^{-4})	-130.93	-133.46	5.04 (5)	0.4105
# of STY HITs	-1.43×10^{-4} (6.07×10^{-4})	-132.48	-133.46	1.95 (5)	0.8563
STY ≤ 50 ppm (ppm-years)	-3.49×10^{-4} (7.54×10^{-4})	-132.53	-133.46	1.85 (5)	0.8690
STY > 50 ppm (ppm-years)	-4.20×10^{-4} (8.60×10^{-4})	-132.61	-133.46	1.69 (5)	0.8905

Thus, the best Cox proportional hazards log linear rate ratio model for CML should adjust for the number of BD HITs in addition to using age as the index variable. However, as indicated at the onset of this analysis, the slope of the log linear rate ratio is not significantly different than zero (either before or after adjustment for these covariates) and the maximum likelihood estimate is negative for the final Cox proportional hazards model that adjusts for statistically significant covariates.

Appendix D

Likelihood-Based Estimation of Exposure Lags for the Poisson Regression Linear Rate Ratio Model and the Cox Proportional Hazards Log Linear Rate Ratio Model of Three Well-Defined Subsets of Leukemia (AML, CLL, and CML)

The standard Poisson regression likelihood function depends heavily on the partition of person years into intervals of covariates (i.e., age groups, calendar year groups, cumulative exposure group, etc.). Because the cumulative exposures are affected by changes in exposure lags, the person years in a specific interval of cumulative exposures are also affected by changes in exposure lags. These changes in the person years assigned to exposure groups makes the standard Poisson regression likelihood non-comparable for different lag periods. The procedure used in this appendix to determine a likelihood-based exposure lag relies on an alternative individualized Poisson regression likelihood function that is more robust to changes in lags. The initial model for Poisson regression (i.e., the linear rate ratio model adjusting for the statistically significant covariates for each of the three subsets of leukemia) with non-lagged exposures is first re-evaluated using the individualized Poisson regression likelihood function and used as a basis for comparison.

The Cox proportional hazards likelihood function does not depend on person years grouped in intervals. Rather, the Cox proportional hazards likelihood function is based on each individual's exposure experience in the cohort and does not present the shortcomings of the standard Poisson regression likelihood. The initial model for Cox proportional (i.e., the log linear rate ratio model adjusting for the statistically significant covariates for each of the three subsets of leukemia) with non-lagged exposures is used as a starting point for comparison.

Exposure Lag for Poisson Regression Model of AML

The maximum likelihood Poisson regression linear rate ratio model for AML adjusts for age and cumulative exposures to DMDTC. This model was fit again using the individualized Poisson regression likelihood function. Using this same likelihood function, the same model was fit to the data with exposures lagged 5, 10, 15, 20, 25, and 30 years. The results are given in Table D.1.

Table D.1. AML: Effect of exposure lags on the maximum likelihood estimation of the Poisson regression linear model on cumulative BD exposure with age and cumulative exposure to DMDTC as categorical covariates

Lag (Years)	Slope of Linear Rate Ratio Model	Individualized Poisson Regression Log Likelihood	Chi Square Statistic	p-value
0	-1.08×10^{-4}	-263.15		n/a
5	-1.09×10^{-4}	-262.49	1.32	0.2506
10	-1.10×10^{-4}	-261.95	2.40	0.1213
15	-1.58×10^{-5}	-263.87	-1.44	n/a
20	5.90×10^{-5}	-263.80	-1.30	n/a
25	-1.76×10^{-4}	-267.56	-8.82	n/a
30	-2.56×10^{-4}	-268.88	-11.46	n/a
35	-4.81×10^{-4}	-269.79	-13.28	n/a

The value of the maximum log likelihood for lags 5 and 10 years were non-significantly larger than the maximum log likelihood with non-lagged exposures. For lags above 10 years, however, the maximum log likelihood values were smaller than the maximum log likelihood for non-lagged exposures. In fact, the model with exposures lagged more than 20 years result in statistically significant reductions in the maximum log likelihood values.

Thus, for the Poisson regression linear rate ratio model for AML, adjusted for age and cumulative exposures to DMDTC, there is no exposure lag with a significantly better likelihood than the same model with non-lagged exposures.

Exposure Lag for Cox Proportional Hazards Model of AML

The maximum likelihood Cox proportional hazards log linear rate ratio model for AML adjusts for cumulative exposures to DMDTC. This model with non-lagged exposures was compared to the same model fit to the data but with exposures lagged 5, 10, 15, 20, 25, and 30 years. The results are given in Table D.2.

Table D.2. AML: Effect of exposure lags on the maximum likelihood estimation of the Cox proportional hazards log-linear model on cumulative BD exposure with cumulative exposure to DMDTC as a categorical covariate

Lag (Years)	Slope of Log-Linear Rate Ratio Model	Log Likelihood	Chi Square Statistic	p-value
0	-2.20×10^{-4}	-217.02		n/a
5	-1.77×10^{-4}	-215.32	3.408	0.0649
10	-1.51×10^{-4}	-218.01	-1.981	n/a
15	-6.37×10^{-5}	-218.89	-3.738	n/a
20	5.91×10^{-5}	-222.15	-10.253	n/a
25	-3.31×10^{-3}	-222.27	-10.501	n/a
30	-3.67×10^{-2}	-218.28	-2.504	n/a
35	-6.08×10^{-2}	-218.61	-3.173	n/a

The value of the maximum log likelihood for lag 5 years was non-significantly larger than the maximum log likelihood with non-lagged exposures. For the model with exposures lagged more than 5 years, however, the maximum log likelihood values were smaller than the maximum log likelihood for the model with non-lagged exposures. In fact, the model with exposures lagged 20 and 25 years result in statistically significant reductions in the maximum log likelihood values.

Thus, for the Cox proportional hazards log linear rate ratio model for AML, adjusted for cumulative exposures to DMDTC, there is no exposure lag with a significantly better likelihood than the same model with non-lagged exposures.

Exposure Lag for Poisson Regression Model of CLL

The maximum likelihood Poisson regression linear rate ratio model for CML adjusts only for age. This model was fit again using the individualized Poisson regression likelihood function. Using this same likelihood function, the same model was fit to the data with exposures lagged 5, 10, 15, 20, 25, and 30 years. The results are given in Table D.3.

Table D.3. CLL: Effect of exposure lags on the maximum likelihood estimation of the Poisson regression linear model on cumulative BD exposure with age as a categorical covariate

Lag (Years)	Slope of Linear Rate Ratio Model	Individualized Poisson Regression Log Likelihood	Chi Square Statistic	p-value
0	2.67×10^{-3}	-249.52		n/a
5	3.08×10^{-3}	-249.24	0.56	0.4543
10	3.63×10^{-3}	-249.13	0.78	0.3771
15	4.37×10^{-3}	-249.14	0.76	0.3833
20	5.27×10^{-3}	-249.52	0.00	1.0000
25	4.07×10^{-3}	-251.42	-3.80	n/a
30	3.76×10^{-3}	-252.14	-5.24	n/a
35	1.10×10^{-2}	-251.54	-4.04	n/a

The values of the maximum log likelihood for the model with exposures lagged 5, 10, and 15 years were non-significantly larger than the maximum log likelihood for the model with non-lagged exposures. For lags above 20 years, however, the maximum log likelihood values were smaller than the maximum log likelihood with non-lagged exposures. In fact, lags longer than 25 years result in statistically significant reductions in the maximum log likelihood values.

Thus, for the Poisson regression linear rate ratio model for CLL, adjusted for age, there is no exposure lag with a significantly better likelihood than the same model with non-lagged exposures.

Exposure Lag for Cox Proportional Hazards Model of CLL

The maximum likelihood Cox proportional hazards log linear rate ratio model for CLL does not adjust for any covariates other than using age as the index variable. This model with non-lagged exposures was compared to the same model fit to the data but with exposures lagged 5, 10, 15, 20, 25, and 30 years. The results are given in Table D.4.

Table D.4. CLL: Effect of exposure lags on the maximum likelihood estimation of the Cox proportional hazards log-linear model on cumulative BD exposure with no other covariates

Lag (Years)	Slope of Log-Linear Rate Ratio Model	Log Likelihood	Chi Square Statistic	p-value
0	4.15×10^{-4}	-205.83		n/a
5	4.53×10^{-4}	-205.58	0.504	0.4777
10	4.71×10^{-4}	-206.03	-0.385	n/a
15	4.95×10^{-4}	-206.51	-1.354	n/a
20	5.57×10^{-4}	-207.19	-2.701	n/a
25	4.78×10^{-4}	-208.15	-4.626	n/a
30	8.89×10^{-4}	-208.04	-4.411	n/a
35	1.82×10^{-3}	-207.91	-4.157	n/a

The value of the maximum log likelihood for lag 5 years was non-significantly larger than the maximum log likelihood for non-lagged exposures. For lags above 5 years, however, the maximum log likelihood values were smaller than the maximum log likelihood for non-lagged exposures. In fact, the model with exposures lagged more than 20 years result in statistically significant reductions in the maximum log likelihood values.

Thus, for the Cox proportional hazards log linear rate ratio model for CLL, there is no exposure lag with a significantly better likelihood than the same model with non-lagged exposures.

Exposure Lag for Poisson Regression Model of CML

The maximum likelihood Poisson regression linear rate ratio model for CML adjusts for age and the number of BD HITs. This model was fit again using the individualized Poisson regression likelihood function. Using this same likelihood function, the same model was fit to the data with exposures lagged 5, 10, 15, 20, 25, and 30 years. The results are given in Table D.5.

Table D.5. CML: Effect of exposure lags on the maximum likelihood estimation of the Poisson regression linear model on cumulative BD exposure with age and cumulative number of BD HITs as categorical covariates

Lag (Years)	Slope of Linear Rate Ratio Model	Individualized Poisson Regression Log Likelihood	Chi Square Statistic	p-value
0	-1.08×10^{-4}	-168.74		n/a
5	-1.09×10^{-4}	-167.74	2.00	0.1573
10	-1.10×10^{-4}	-166.99	3.50	0.0614
15	-1.11×10^{-4}	-162.05	13.38	**0.0003****
20	-1.34×10^{-4}	-169.73	-1.98	n/a
25	-1.76×10^{-4}	-174.07	-10.66	n/a
30	-2.57×10^{-4}	-175.10	-12.72	n/a
35	-4.81×10^{-4}	-176.17	-14.86	n/a

**statistically significant at the 1% significance level

The values of the maximum log likelihood for lags 5, 10 and 15 years were larger than the maximum log likelihood for non-lagged exposures. In addition, the model fit with exposures lagged 15 years resulted in a statistically significant larger maximum log likelihood value than the model fit with non-lagged exposures. For lags above 15 years, however, the maximum log likelihood values were smaller than the maximum log likelihood for non-lagged exposures. In fact, the model with exposures lagged longer than 20 years result in statistically significant reductions in the maximum log likelihood values.

Thus, for the Poisson regression linear rate ratio model for CML, adjusted for age and cumulative number of BD HITs, the model with exposures lagged 15 years fits the observed data statistically significantly better than the same model with non-lagged exposures.

Exposure Lag for Cox Proportional Hazards Model of CML

The maximum likelihood Cox proportional hazards log linear rate ratio model for CML adjusts for cumulative number of BD HITs. This model with non-lagged exposures was compared to the same model fit to the data but with exposures lagged 5, 10, 15, 20, 25, and 30 years. The results are given in Table D.6.

Table D.6. CML: Effect of exposure lags on the maximum likelihood estimation of the Cox proportional hazards log-linear model on cumulative BD exposure with cumulative number of BD HITs as a categorical covariate

Lag (Years)	Slope of Log-Linear Rate Ratio Model	Log Likelihood	Chi Square Statistic	p-value
0	-1.29×10^{-4}	-133.46		n/a
5	-1.81×10^{-4}	-132.96	0.998	0.3178
10	-1.33×10^{-4}	-133.69	-0.476	n/a
15	4.11×10^{-5}	-129.51	7.900	**0.0049****
20	8.71×10^{-5}	-135.18	-3.457	n/a
25	-3.40×10^{-4}	-137.90	-8.897	n/a
30	-3.08×10^{-3}	-139.45	-11.996	n/a
35	-4.87×10^{-3}	-139.49	-12.066	n/a

**statistically significant at the 1% significance level

The values of the maximum log likelihood for lags 5 and 15 years were larger than the maximum log likelihood for non-lagged exposures. In addition, the model fit with exposures lagged 15 years resulted in a statistically significant larger maximum log likelihood value than the model fit with non-lagged exposures. For lags of 10 years and lags above 15 years, however, the maximum log likelihood values were smaller than the maximum log likelihood for non-lagged exposures. In fact, the model with exposures lagged longer than 20 years result in statistically significant reductions in the maximum log likelihood values.

Thus, for the Cox proportional hazards log linear rate ratio model for CML, adjusted for cumulative number of BD HITs, the model with exposures lagged 15 years fits the observed data statistically significantly better than the same model with non-lagged exposures.

Appendix E

Actuarial Computation of Lifetime Added Risks of AML, CLL and CML for Occupational Exposures Using Population-Based Age-Specific Background Rates and Competing Risks:

SCOEL Methodology as described in Risk Assessment for 1,3-Butadiene; Scientific Committee Group on Occupational Exposure Limits; SCOEL/INF/521, Luxembourg, June 2002; Received from Prof. A. Bertazzi

This appendix lists the added risks of AML, CLL, and CML calculated for with the final models adjusting for the statistically significant covariates and the statistically significant lag periods. European-specific age-adjusted background mortality rates for AML, CLL and CML along with age-adjusted survival probabilities are used to calculate added risks applying SCOEL's methodology. The lifetime added risks are calculated at age 85 years for non-zero occupational exposures to BD during 45 potential years of work with the person being exposed between 20 and 65 years of age and ignoring exposures received 40 or more years ago and also excluding exposures received within the lag period. It is assumed that there are no tasks involving exposure to BD concentrations higher than 100 ppm (i.e., no BD HITs) and that there are no other non-background exposures that could contribute to the development of AML, CLL, or CML. The BD exposure concentration is specified to be the same for each year (0.1, 0.2, 0.5, 1, 2, 5, or 10 ppm (the occupational exposure concentrations used by SCOEL)).

Table E.1, E.4, and E.7 show the added risks based on the maximum likelihood estimates and on the 95% upper confidence limits on the slopes of the final Poisson and Cox models for AML, CLL, and CML, respectively. Tables E.2, E.5, and E.8 show the ratio of the upper confidence limit on the added risks to the MLE of the added risks of the Poisson and Cox models for AML, CLL, and CML, respectively. Tables E.3, E.6, and E.9 list the ratios of the added risks based on the Poisson linear rate ratio model to the added risks based on the Cox proportional hazards linear model for AML, CLL, and CML, respectively.

SCOEL Methodology

In the August 28, 2006, report "Quantitative Risk Assessment of Exposures to 1,3-Butadiene in EU Occupational Settings Based on the University of Alabama at Birmingham Epidemiological Study" by Robert L. Sielken Jr. and Ciriaco Valdez-Flores, Ph.D., the final model for all leukemia from the ACC report was coupled with the Scientific Committee Group on Occupational Exposure Limits (SCOEL) methodology for the actuarial computation of added risks for all leukemia for rate ratio models using European Union (EU) population-based age-specific background rates and competing risks. For that report, the SCOEL methodology was implemented in a spreadsheet format in Microsoft Excel 2000. The three examples in the SCOEL reference were duplicated as evidence of the correct implementation of the spreadsheet. These were verification examples and not the specific values generated for the butadiene lifetime added risks calculated for this report.

In the SCOEL reference, it is noted that "A life-table for European population does not exist and thus a national reference had to be used to take into account the natural decline of the population by age. For this exercise the population of England & Wales has been adopted, for at least three reasons: among the European populations it has a long standing tradition in vital statistics; the data are easily available; the 1981 life-table has been used in similar exercises." (page 3).

In this report (as in the August 28, 2006, report), the 1981 rates for England & Wales are used. The rates for all leukemia and all cause mortality (survival probabilities) are exactly the same as in the SCOEL reference. The rates for AML, CLL, and CML are the 1981 rates for England & Wales. These rates are listed in Table E.10.

As a sensitivity analysis, the added risk calculations were all repeated with age-specific rates of each of the subtypes of leukemia replaced by a moving average of 3 age-specific rates (the age interval before, the age interval itself, and the age interval after). The added risks calculated using these smoothed age-specific rates were virtually unchanged compared to the added risks calculated using non-smoothed age-specific rates. This suggests that the variability in the 1981 age-specific rates for subtypes of leukemia due to the decreased frequency of these subtypes compared to that for all leukemia does not substantially impact the calculation of the added risks herein for these subtypes of leukemia. In other words, the calulations of the added risks for subtypes of leukemia in this report are fairly robust to the estimates of the background rates of these subtypes of leukemia.

Table E.1. Lifetime added risks of AML per 1,000 and three alternative definitions of AML for a grid of occupational BD exposure concentrations using the maximum likelihood estimate of the Poisson regression linear model and the Cox proportional hazards log linear model

BD ppm	Model	Leukemia subtypes included with the primary subset			
		Primary	Primary + Other/Unknown	Primary + Unspecified	Primary + Unspecified + Other/Unknown
		Based on the Maximum Likelihood Estimates of the Slopes			
0.1	Poisson	0[1]	0	0	0
	Cox	0	0	0	0
0.2	Poisson	0	0	0	0
	Cox	0	0	0	0
0.5	Poisson	0	0	0	0
	Cox	0	0	0	0
1.0	Poisson	0	0	0	0
	Cox	0	0	0	0
2.0	Poisson	0	0	0	0
	Cox	0	0	0	0
5.0	Poisson	0	0	0	0
	Cox	0	0	0	0
10.0	Poisson	0	0	0	0
	Cox	0	0	0	0
		Based on the 95% Upper Confidence Limit on the Estimated Slopes			
0.1	Poisson	0.00140	0.00329	0.00445	0.00616
	Cox	0.00414	0.00364	0.00397	0.00361
0.2	Poisson	0.00279	0.00659	0.00889	0.01232
	Cox	0.00829	0.00729	0.00795	0.00724
0.5	Poisson	0.00698	0.01646	0.02223	0.03080
	Cox	0.02081	0.01828	0.01993	0.01814
1.0	Poisson	0.01396	0.03293	0.04447	0.06161
	Cox	0.04186	0.03674	0.04007	0.03644
2.0	Poisson	0.02793	0.06585	0.08893	0.12321
	Cox	0.08469	0.07419	0.08099	0.07356
5.0	Poisson	0.06982	0.16462	0.22230	0.30798
	Cox	0.21933	0.19103	0.20908	0.18911
10.0	Poisson	0.13962	0.32921	0.44453	0.61583
	Cox	0.46568	0.40158	0.44150	0.39645

[1]an added risk equal to 0 indicates that the slope of the model was non-positive, resulting in a dose-response relationship that would not estimate a positive added risk for any positive exposure to BD

Table E.2. Ratio of the upper confidence limit of the lifetime added risks of AML per 1,000 to the maximum likelihood estimate of the lifetime added risk of AML per 1,000 and three alternative definitions of AML for a grid of occupational BD exposure concentrations using the maximum likelihood estimate of the Poisson regression linear model and the Cox proportional hazards log linear model

BD ppm	Model	Leukemia subtypes included with the primary subset			
		Primary	Primary + Other/Unknown	Primary + Unspecified	Primary + Unspecified + Other/Unknown
0.1	Poisson	inf[1]	inf	inf	inf
	Cox	inf	inf	inf	inf
0.2	Poisson	inf	inf	inf	inf
	Cox	inf	inf	inf	inf
0.5	Poisson	inf	inf	inf	inf
	Cox	inf	inf	inf	inf
1.0	Poisson	inf	inf	inf	inf
	Cox	inf	inf	inf	inf
2.0	Poisson	inf	inf	inf	inf
	Cox	inf	inf	inf	inf
5.0	Poisson	inf	inf	inf	inf
	Cox	inf	inf	inf	inf
10.0	Poisson	inf	inf	inf	inf
	Cox	inf	inf	inf	inf

[1] inf indicates that the ratio cannot be calculated because the upper confidence limit on added risk is greater than zero and the maximum likelihood estimate of the added risk is equal to zero resulting in a value equal to infinity

Table E.3. Ratio of the lifetime added risks of AML per 1,000 based on the the Poisson regression linear model to the lifetime added risk of AML per 1,000 based on the Cox proportional hazards log linear model for a grid of occupational BD exposure concentrations and three alternative definitions of AML

BD ppm	Leukemia subtypes included with the primary subset			
	Primary	Primary + Other/Unknown	Primary + Unspecified	Primary + Unspecified + Other/Unknown
Ratios of Added Risks Based on the Maximum Likelihood Estimates of the Slopes				
0.1	n/a[1]	n/a	n/a	n/a
0.2	n/a	n/a	n/a	n/a
0.5	n/a	n/a	n/a	n/a
1.0	n/a	n/a	n/a	n/a
2.0	n/a	n/a	n/a	n/a
5.0	n/a	n/a	n/a	n/a
10.0	n/a	n/a	n/a	n/a
Ratios of Added Risks Based on the 95% Upper Confidence Limit on the Slopes				
0.1	0.3	0.9	1.1	1.7
0.2	0.3	0.9	1.1	1.7
0.5	0.3	0.9	1.1	1.7
1.0	0.3	0.9	1.1	1.7
2.0	0.3	0.9	1.1	1.7
5.0	0.3	0.9	1.1	1.6
10.0	0.3	0.8	1.0	1.6

[1] n/a indicates that the ratio cannot be calculated because added risks based on the Poisson regression model and on the Cox proportional hazards model are both zero

Table E.4. Lifetime added risks of CLL per 1,000 and three alternative definitions of CLL for a grid of occupational BD exposure concentrations using the maximum likelihood estimate of the Poisson regression linear model and the Cox proportional hazards log linear model

BD ppm	Model	Leukemia subtypes included with the primary subset			
		Primary	Primary + Other/Unknown	Primary + Unspecified	Primary + Unspecified + Other/Unknown
		Based on the Maximum Likelihood Estimates of the Slopes			
0.1	Poisson	0.01104	0.01293	0.00897	0.01065
	Cox	0.00161	0.00160	0.00164	0.00163
0.2	Poisson	0.02207	0.02587	0.01795	0.02131
	Cox	0.00322	0.00320	0.00329	0.00327
0.5	Poisson	0.05518	0.06467	0.04487	0.05327
	Cox	0.00806	0.00800	0.00824	0.00818
1.0	Poisson	0.11036	0.12934	0.08973	0.10653
	Cox	0.01618	0.01606	0.01654	0.01642
2.0	Poisson	0.22071	0.25865	0.17946	0.21305
	Cox	0.03258	0.03234	0.03330	0.03305
5.0	Poisson	0.55164	0.64643	0.44855	0.53249
	Cox	0.08315	0.08253	0.08491	0.08426
10.0	Poisson	1.10281	1.29223	0.89680	1.06456
	Cox	0.17214	0.17081	0.17558	0.17418
		Based on the 95% Upper Confidence Limit on the Estimated Slopes			
0.1	Poisson	0.02429	0.02739	0.02016	0.02285
	Cox	0.00245	0.00241	0.00254	0.00251
0.2	Poisson	0.04857	0.05479	0.04032	0.04571
	Cox	0.00491	0.00483	0.00509	0.00502
0.5	Poisson	0.12142	0.13696	0.10081	0.11426
	Cox	0.01230	0.01212	0.01276	0.01260
1.0	Poisson	0.24283	0.27389	0.20160	0.22850
	Cox	0.02474	0.02436	0.02566	0.02532
2.0	Poisson	0.48556	0.54767	0.40313	0.45692
	Cox	0.04999	0.04921	0.05184	0.05116
5.0	Poisson	1.21324	1.36833	1.00737	1.14171
	Cox	0.12895	0.12690	0.13364	0.13184
10.0	Poisson	2.42424	2.73382	2.01320	2.28144
	Cox	0.27198	0.26742	0.28156	0.27758

Table E.5. Ratio of the upper confidence limit of the lifetime added risks of CLL per 1,000 to the maximum likelihood estimate of the lifetime added risk of CLL per 1,000 and three alternative definitions of CLL for a grid of occupational BD exposure concentrations using the maximum likelihood estimate of the Poisson regression linear model and the Cox proportional hazards log linear model

BD ppm	Model	Leukemia subtypes included with the primary subset			
		Primary	Primary + Other/Unknown	Primary + Unspecified	Primary + Unspecified + Other/Unknown
0.1	Poisson	2.2	2.1	2.2	2.1
	Cox	1.5	1.5	1.5	1.5
0.2	Poisson	2.2	2.1	2.2	2.1
	Cox	1.5	1.5	1.5	1.5
0.5	Poisson	2.2	2.1	2.2	2.1
	Cox	1.5	1.5	1.5	1.5
1.0	Poisson	2.2	2.1	2.2	2.1
	Cox	1.5	1.5	1.6	1.5
2.0	Poisson	2.2	2.1	2.2	2.1
	Cox	1.5	1.5	1.6	1.5
5.0	Poisson	2.2	2.1	2.2	2.1
	Cox	1.6	1.5	1.6	1.6
10.0	Poisson	2.2	2.1	2.2	2.1
	Cox	1.6	1.6	1.6	1.6

Table E.6. Ratio of the lifetime added risks of CLL per 1,000 based on the the Poisson regression linear model to the lifetime added risk of CLL per 1,000 based on the Cox proportional hazards log linear model for a grid of occupational BD exposure concentrations and three alternative definitions of CLL

BD ppm	Leukemia subtypes included with the primary subset			
	Primary	Primary + Other/Unknown	Primary + Unspecified	Primary + Unspecified + Other/Unknown
Ratios of Added Risks Based on the Maximum Likelihood Estimates of the Slopes				
0.1	6.9	8.1	5.5	6.5
0.2	6.9	8.1	5.5	6.5
0.5	6.8	8.1	5.4	6.5
1.0	6.8	8.1	5.4	6.5
2.0	6.8	8.0	5.4	6.4
5.0	6.6	7.8	5.3	6.3
10.0	6.4	7.6	5.1	6.1
Ratios of Added Risks Based on the 95% Upper Confidence Limit on the Slopes				
0.1	9.9	11.4	7.9	9.1
0.2	9.9	11.3	7.9	9.1
0.5	9.9	11.3	7.9	9.1
1.0	9.8	11.2	7.9	9.0
2.0	9.7	11.1	7.8	8.9
5.0	9.4	10.8	7.5	8.7
10.0	8.9	10.2	7.2	8.2

Table E.7. Lifetime added risks of CML per 1,000 and three alternative definitions of CML for a grid of occupational BD exposure concentrations using the maximum likelihood estimate of the Poisson regression linear model and the Cox proportional hazards log linear model

BD ppm	Model	Leukemia subtypes included with the primary subset			
		Primary	Primary + Other/Unknown	Primary + Unspecified	Primary + Unspecified + Other/Unknown
		Based on the Maximum Likelihood Estimates of the Slopes			
0.1	Poisson	0[1]	0	0.00033	0.00041
	Cox	0.00007	0.00016	0.00017	0.00025
0.2	Poisson	0	0	0.00067	0.00082
	Cox	0.00014	0.00033	0.00034	0.00051
0.5	Poisson	0	0	0.00167	0.00205
	Cox	0.00034	0.00082	0.00086	0.00127
1.0	Poisson	0	0	0.00334	0.00410
	Cox	0.00069	0.00164	0.00172	0.00254
2.0	Poisson	0	0	0.00668	0.00820
	Cox	0.00138	0.00329	0.00344	0.00509
5.0	Poisson	0	0	0.01669	0.02051
	Cox	0.00345	0.00825	0.00862	0.01279
10.0	Poisson	0	0	0.03339	0.04102
	Cox	0.00691	0.01660	0.01733	0.02577
		Based on the 95% Upper Confidence Limit on the Estimated Slopes			
0.1	Poisson	0.00114	0.00125	0.00288	0.00311
	Cox	0.00206	0.00188	0.00207	0.00195
0.2	Poisson	0.00228	0.00250	0.00575	0.00622
	Cox	0.00413	0.00376	0.00415	0.00391
0.5	Poisson	0.00570	0.00624	0.01438	0.01554
	Cox	0.01036	0.00943	0.01042	0.00982
1.0	Poisson	0.01139	0.01248	0.02876	0.03109
	Cox	0.02087	0.01899	0.02096	0.01975
2.0	Poisson	0.02278	0.02496	0.05751	0.06217
	Cox	0.04233	0.03847	0.04246	0.03997
5.0	Poisson	0.05695	0.06240	0.14377	0.15542
	Cox	0.11052	0.10004	0.11031	0.10361
10.0	Poisson	0.11390	0.12479	0.28750	0.31081
	Cox	0.23791	0.21392	0.23548	0.22032

[1]an added risk equal to 0 indicates that the slope of the model was non-positive, resulting in a dose-response relationship that would not estimate a positive added risk for any positive exposure to BD

Table E.8. Ratio of the upper confidence limit of the lifetime added risks of CML per 1,000 to the maximum likelihood estimate of the lifetime added risk of CML per 1,000 and three alternative definitions of CML for a grid of occupational BD exposure concentrations using the maximum likelihood estimate of the Poisson regression linear model and the Cox proportional hazards log linear model

BD ppm	Model	Leukemia subtypes included with the primary subset			
		Primary	Primary + Other/Unknown	Primary + Unspecified	Primary + Unspecified + Other/Unknown
0.1	Poisson	inf[1]	inf	8.7	7.6
	Cox	29.4	11.8	12.2	7.8
0.2	Poisson	inf	inf	8.6	7.6
	Cox	29.5	11.4	12.2	7.7
0.5	Poisson	inf	inf	8.6	7.6
	Cox	30.5	11.5	12.1	7.7
1.0	Poisson	inf	inf	8.6	7.6
	Cox	30.2	11.6	12.2	7.8
2.0	Poisson	inf	inf	8.6	7.6
	Cox	30.7	11.7	12.3	7.9
5.0	Poisson	inf	inf	8.6	7.6
	Cox	32.0	12.1	12.8	8.1
10.0	Poisson	inf	inf	8.6	7.6
	Cox	34.4	12.9	13.6	8.5

[1]inf indicates that the ratio cannot be calculated because the upper confidence limit on added risk is greater than zero and the maximum likelihood estimate of the added risk is equal to zero resulting in a value equal to infinity

Table E.9. Ratio of the lifetime added risks of AML per 1,000 based on the the Poisson regression linear model to the lifetime added risk of AML per 1,000 based on the Cox proportional hazards log linear model for a grid of occupational BD exposure concentrations and three alternative definitions of AML

BD ppm	Leukemia subtypes included with the primary subset			
	Primary	Primary + Other/Unknown	Primary + Unspecified	Primary + Unspecified + Other/Unknown
Ratios of Added Risks Based on the Maximum Likelihood Estimates of the Slopes				
0.1	0.0	0.0	1.9	1.6
0.2	0.0	0.0	2.0	1.6
0.5	0.0	0.0	1.9	1.6
1.0	0.0	0.0	1.9	1.6
2.0	0.0	0.0	1.9	1.6
5.0	0.0	0.0	1.9	1.6
10.0	0.0	0.0	1.9	1.6
Ratios of Added Risks Based on the 95% Upper Confidence Limit on the Slopes				
0.1	0.6	0.7	1.4	1.6
0.2	0.6	0.7	1.4	1.6
0.5	0.6	0.7	1.4	1.6
1.0	0.5	0.7	1.4	1.6
2.0	0.5	0.6	1.4	1.6
5.0	0.5	0.6	1.3	1.5
10.0	0.5	0.6	1.2	1.4

Table E.10. The European-specific age-adjusted background mortality rates for AML, CLL, and CML along with age-adjusted survival probabilities that were used to calculate added risks by applying SCOEL's methodology:

Year	Sex	Age	Deaths in General Population	
			All Causes	All Leukemia
1981	M	20-24	1576	31
1981	M	25-29	1427	37
1981	M	30-34	1754	42
1981	M	35-39	2143	25
1981	M	40-44	3392	41
1981	M	45-49	5920	62
1981	M	50-54	10969	73
1981	M	55-59	19688	133
1981	M	60-64	27170	145
1981	M	65-69	40298	243
1981	M	70-74	51891	273
1981	M	75-79	50959	255
1981	M	80-84	35815	177

Age	Deaths in General Population			
	AML Only: 205.0, 206.0	AML+Acute Leukemia 205.0, 206.0 plus 207.0	AML+Unspecified Myelogenous: 205.0, 206.0 plus 205.9	AML+Acute Leukemia+Unspecified Myelogenous: 205.0, 206.0 plus 207.0 plus 205.9
20-24	12	12	13	13
25-29	18	19	18	19
30-34	22	23	22	23
35-39	10	11	10	11
40-44	24	24	24	24
45-49	25	27	25	27
50-54	33	34	33	34
55-59	59	62	62	65
60-64	53	54	58	59
65-69	93	98	100	105
70-74	108	113	110	115
75-79	71	76	78	83
80-84	59	62	63	66

Table E.10. (Continued)

Age	Deaths in General Population			
	CLL Only: 204.1	CLL + Chronic Leukemia: 204.1 plus 207.1	CLL + Unspecified Lymphocytic: 204.1 plus 204.9	CLL + Chronic Leukemia + Unspecified Lymphocytic: 204.1 plus 207.1 plus 204.9
20-24	0	0	0	0
25-29	0	0	0	0
30-34	1	1	1	1
35-39	0	0	0	0
40-44	1	1	1	1
45-49	4	4	4	4
50-54	10	10	11	11
55-59	23	23	27	27
60-64	37	37	38	38
65-69	68	68	71	71
70-74	79	79	84	84
75-79	83	83	86	86
80-84	55	55	59	59

Age	Deaths in General Population			
	CML Only: 205.1	CML + Chronic Leukemia: 205.1 plus 207.1	CML + Unspecified Myelogenous: 205.1 plus 205.9	CML + Chronic Leukemia + Unspecified Myelogenous: 205.1 plus 207.1 plus 205.9
20-24	6	6	7	7
25-29	7	7	7	7
30-34	9	9	9	9
35-39	9	9	9	9
40-44	14	14	14	14
45-49	20	20	20	20
50-54	15	15	15	15
55-59	23	23	26	26
60-64	16	16	21	21
65-69	32	32	39	39
70-74	37	37	39	39
75-79	43	43	50	50
80-84	24	24	28	28