MEMORANDUM

To: Judy Facey, EPA; Alexander Kliminsky, EPA; Lori Brunsman, EPA; Aaron Niman, EPA; Matthew Crowley, EPA; Deborah Burgin, EPA

From: Sorina Eftim, Jenna Spruce, Courtney Rosenthal, Ryan Gan, ICF

Date: February 21, 2023

Re: Statistical Reanalysis of Data from Two Formaldehyde Inhalation Exposure Studies: Lang et al. (2008) and Mueller et al. (2013).

1. Introduction

This memorandum describes ICF's reanalysis of the summary data provided in the Lang et al. (2008) and Mueller et al. (2013) formaldehyde inhalation exposure studies. Detailed descriptions of the proposed analyses and methods were provided in our previous memo ("Notes on formaldehyde papers by Lang et al. 2008 and Mueller et al. 2013") delivered on November 4, 2022 (Appendix A.). EPA's attempts to obtain the raw data from the study authors was unsuccessful. Therefore, the analyses presented here used summary data available in figures, tables, and supplemental material from the two studies. This memorandum is accompanied by an Excel file (Appendix 2) and the code used in and output files from the statistical analyses (Appendix 3). Appendix 4 includes BMDS modelling outputs.

The memorandum presents a brief background on the two studies evaluated, the statistical methods used to evaluate specific endpoints, and the results of these analyses. As illustrated in **Table 1** the results of the reanalysis presented are mixed. We were able to confirm certain findings, but none for blinking frequency and reaction times from Lang et al. (2008). Except for blinking frequency, BMD analyses resulted either in no valid models, or in models yielding BMDs and BMDLs higher than the maximum concentration, suggesting caution in using these endpoints for POD derivation.

2. Background

Lang et al. (2008) and Mueller et al. (2013) are two controlled human exposure studies (short term) reviewed in the recent Toxicological Review of Formaldehyde – Inhalation (EPA, 2022a) that evaluated sensory irritation in humans. The two studies were eligible but not selected for point of departure (POD) derivation because "difficult to define adverse level cut-off for these endpoints" (eye irritation: conjunctival redness, blinking frequency, symptom score) for Lang et al. (2008); and because "an exposure-response trend was not observed for either endpoint. Difficult to define an adverse response

level cutoff for these endpoints" for Mueller et al. (2013). The endpoints considered were eye irritation, measured by the tear film break-up time and by the symptom score using visual analogue scale (VAS).

Briefly, Lang et al. (2008) measured the effects of inhaled formaldehyde vapor on conjunctival redness, blinking frequency, nasal flow and resistance, pulmonary function, reaction times to acoustic and/or visual stimuli, and subjective ratings of eye irritation, nasal irritation, olfactory symptoms, and respiratory irritation. Twenty-one healthy non-smoking human subjects (11 males, 10 females, age range 19 to 39 years) were examined in groups of four at a time. Each subject served as their own control. Each subject was exposed for 4 h to each of the 10 exposure conditions on 10 consecutive working days. Exposure order was randomly assigned and double-blinded. Air concentrations of 0, 0.15, 0.3, and 0.5 ppm (0.0, 0.19, 0.37, 0.62 mg/m3 formaldehyde) with additional concentrations of 0.3 and 0.5 ppm with peaks up to 1.0 ppm (1.23 mg/m3 formaldehyde) were used in this study. Additional formaldehyde concentrations of 0.0, 0.3, 0.5 ppm with ethyl acetate (EA) to mask formaldehyde odor were also used. Ventilators assured homogenous distribution of the formaldehyde and EA test chamber atmosphere generated under quasi-static conditions. Baseline measurements of nasal resistance and flow (using rhinomanometry) and subjective ratings of physical symptoms (using the SPES questionnaire) were collected each day before exposure. Cycle ergometry, the SPES questionnaire, digital slit lamp photography to measure conjunctival redness, and video recording of blinking frequency were administered at the start of exposure and at 120 and 195 min of exposure. Post-exposure assessments were the SPES questionnaire, measures of nasal resistance and flow, and reaction time measurements that were done immediately after the end of exposure. On the final day after exposure ended, a physical examination and assessment of pulmonary function were done. Formaldehyde concentrations in the chamber were measured by collection of two air samples taken each exposure day and analyzed using dinitrophenylhydrazine and HPLC analysis. Analytical concentrations were not reported.

Mueller et al. (2013) measured the effects of inhaled formaldehyde vapor on conjunctival redness, blinking frequency, tear film break-up time, nasal resistance and flow, subjective ratings of eye irritation, nasal irritation, olfactory symptoms, and other subjective symptoms. Forty-one healthy non-smoking human subjects (all male, average age 32 years) participated in this study and were examined in groups of two. Subjects were categorized as hyposensitive (20 subjects) or hypersensitive (21 subjects) based on CO_2 sensitivity measurements in nasal mucosa. Each subject served as their own control. Each subject received 5 four-hour exposures to formaldehyde or control air over 5 consecutive days. Exposure order was randomly assigned. Air concentrations were 0 ppm, 0.3 ppm + 4 peaks of 0.6 ppm, 0.4 ppm + 4 peaks of 0.8 ppm, 0.5 ppm and 0.7 ppm (0.0, 0.37 + 0.74, 0.49 + 0.98, 0.62, and 0.86 mg/m^3). Peak exposures were 15 minutes each. Formaldehyde concentrations in the chamber were measured by real-time monitoring and by HPLC analysis of baseline and peak exposures. Baseline measurements for nasal flow rates, self-reported tear film break-up time, CO_2 sensitivity, conjunctival redness, eye-blinking frequency, and subjective symptoms were collected within 1 hour before start of exposure each day. During the exposure, subjects performed four cycle ergometer units at 80 watts for 15 min at predefined times; two of the four ergometric units were carried out during an exposure peak for days with exposure peaks. Post-exposure recording of eye-blinking frequency and completion of the SPES questionnaire took place during the last 15 minutes of exposure, while all other outcome measures were examined within 1 hour after the end of exposure each day.

ICF was tasked with reproducing and confirming results in the two studies. Appendix A describes the approaches proposed assuming presence or absence of the original (raw) data from each study. The following section describe the statistical methods used in the analyses that could be performed using only the summary data reported in the study publication. **Table 1** outlines briefly which analyses were performed and those that could not be performed in absence of the raw data, or additional detail from study authors.

Table 1. Summary of study-specific endpoints with analyses using summary data and endpoints without ICF analyses.

Endpoint	Published result	ICF Result
Lang et al. (2008)		
Endpoints with analyses using summary data		
Conjunctival redness	Conjunctival redness significantly increased at 0.5 ppm with peaks of 1.0 ppm (without EA) compared to the 0-ppm control condition	**Confirmed finding of** significantly different conjunctival redness rates across the different doses only 195 minutes after exposure. There was no difference in eye redness rates at 15 or 120 minutes after exposure (p-values of 0.367, and 0.216, respectively. BMDs (BMDLs) of 0.49 (0.18) ppm at 15 minutes after exposure for concentrations without EA, and 0.23 (0.08) ppm at 120 minutes after exposure for doses with EA.
Blinking frequency	Blinking frequency significantly increased at 0.5 ppm with 1.0 ppm peaks with or without EA.	No statistically significant differences were observed. BMDs and BMDLs higher than the maximum concentration, suggesting caution in using this approach.
Reaction times	Visual stimulus reaction time increased significantly at 0.3 ppm with or without EA. Acoustic stimulus reaction time increased significantly at 0.3 ppm without EA. Combined visual and acoustic reaction time increased significantly at 0.3 ppm without EA.	No statistically significant differences were observed. BMD approaches resulted either in no valid models, or in BMDs and BMDLs higher than the maximum concentration.
SPES total score	Mean total symptom score was significantly higher versus the controls at 0.5 ppm with peaks, with or without EA.	**Confirmed in ICF analysis.** Compared to the 0 ppm-control group, significant increases in total symptom scores were observed at all concentrations except at 0.15 ppm and 0.3 ppm without EA. Compared to the 0 ppm with EA co-exposure group, mean total symptom score was significantly higher only at concentrations of 0 ppm and 0.15 ppm FA

Endpoint	Published result	ICF Result
		without EA co-exposure. However, the results need to be interpreted with extreme caution give the clear violation of normality illustrated in Figure 5. BMDs and BMDLs higher than the maximum concentration, suggesting caution in using this approach.
SPES eye irritation	In most cases the mean scores were significantly higher than the controls.	**Confirmed some findings**. Compared to the 0 ppm-control group, significant increases in eye irritation scores were observed at concentrations of 0.3 ppm and 0.5 ppm with peaks without EA co-exposure, and at 0.5 ppm with or without peaks, with EA co-exposure. Compared to the 0 ppm + EA co-exposure, mean eye irritation score was significantly higher only at concentrations of 0.5 ppm with four 1.0 ppm peaks with or without EA co-exposure. BMDs and BMDLs higher than the maximum concentration, suggesting caution in using this approach.
SPES nasal irritation	Significantly higher scores at 0.5 ppm with peaks of 1.0 ppm with or without EA compared to either control, and at 0.3 ppm + EA or 0.5 ppm + EA compared to 0 ppm-control.	Confirmed the findings. In addition, a significantly higher mean score was observed at 0.15 ppm without EA compared to 0 ppm + EA. BMDs and BMDLs higher than the maximum concentration, suggesting caution in using this approach.
SPES olfactory symptoms	Significantly increased scores at 0.3 ppm FA and higher, and peak exposure and co-exposure with EA increased scores. In addition, only at 0.5 ppm with peaks and EA scores were significantly higher than the 0 ppm +EA control. confirmed the reported findings, except for a lack of significant differences at 0.5 ppm compared to 0 ppm, and at 0.5 ppm with peaks and EA compared to 0 ppm +EA.	**Confirmed the reported findings**, **except** for a lack of significant differences at 0.5 ppm compared to 0 ppm, and at 0.5 ppm with peaks and EA compared to 0 ppm +EA. For concentrations without EA, the estimated BMDs is 1.00 ppm with a BMDL of 0.51 ppm, both higher than the maximum concentration suggesting caution in using such result for POD derivation. No viable model converged at concentrations with EA co-exposure.
SPES respiratory symptoms	Significantly increased scores at 0.3 ppm FA without peaks, at 0.5 ppm with peaks, and at all levels with EA co-exposures.	**Confirmed only a single finding**, of significant increase in mean score at 0.5 ppm with peaks and EA co-exposure compared to the 0 ppm-control.

Endpoint	Published result	ICF Result
		For concentrations without EA, the estimated BMD is 0.78 ppm (higher than the maximum concentration), with a BMDL of 0.35 ppm. For concentrations with EA co-exposure, the estimated BMD is 3.04 ppm with a BMDL of 0.64 ppm, both higher than the maximum concentration suggesting caution in using such result for POD derivation.
SPES annoyance ratings	Increased with elevated concentrations, and at all levels with co-exposure to EA, including the 0 ppm + EA control	**Confirmed only a single finding**, of significant increase in mean score at 0.3 ppm with EA co-exposure compared to the 0 ppm-control. For concentrations without EA, the estimated BMD is 0.52 ppm, with a BMDL of 0.51 ppm. For concentrations with EA co-exposure, the estimated BMD is 2.82 ppm with a BMDL of 0.73. The models yielded BMDs and BMDLs higher than the maximum concentration suggesting caution in using such result for POD derivation.
Correlation of blinking frequency with eye irritation	Correlation of blinking frequency with eye irritation at 0.5 ppm with peaks, with EA (rho=0.54, p=0.01) or without EA (rho=0.36, p=0.10).	**Confirmed findings.** Spearman rank correlations of 0.51 (p-value 0.02) and 0.36 (p-value =0.11).
Endpoints without ICF Analyses		
Nasal resistance and flow	Differences after exposure compared to before exposure were not statistically significant and occurred for the control doses, and thus were assumed to be unrelated to the formaldehyde or EA exposure (data not shown).	Without data, no statistical analyses are feasible. ICF was unable to verify the results for these outcomes.
Pulmonary function	Differences between day 1 and day 10 were not statistically significant (data not shown).	
Mueller et al. (2013)		
Endpoints with analyses using summary data		
Eye-blinking frequency	Statistically significant change for hypersensitives at 0 ppm. In general, decreases after exposure to FA, but no consistent statistically significant change. Differences between hypo- and hyper-sensitives were not statistically significant based on ANCOVA. A	**Confirmed findings.** Statistically significant change for hypersensitives at a 0 ppm. No model was viable for the data in hyposensitives. The estimated BMD and BMDLs in hypersensitives are 8.38 ppm and 0.72 ppm, respectively, higher than the maximum concentration suggesting caution in using such result for POD derivation.

Endpoint	Published result	ICF Result
	consistent dose-effect relationship was not found.	
Tear film break-up time	Significantly increased in 0.4/0.8 ppm and 0.5 ppm ($p < 0.05$).	**Confirmed findings** of significant mean changes after exposure compared to pre-exposure at 0.4 ppm with peaks in both hyper- and hyposensitives, and at 0.5 ppm in hypersensitives. Statistically significantly decreases were also observed in hypersensitives when compared to hyposensitives at 0 ppm. Did not confirm the reported significant changes in hyposensitives at 0.3 ppm with peaks and at 0.7 ppm compared with the 0 ppm, but significant differences were observed in hypersensitives at 0.4 ppm with peaks and at 0.5 ppm when compared to 0 ppm. Models in hyposensitives and hypersensitives yielded BMDs and BMDLs higher than the maximum concentration suggesting caution in using such result for POD derivation.
Nasal flow	Statistically significant change in the mean for hypersensitives at 0.7 ppm. Statistically significant difference in the responses of hyposensitives at 0 ppm versus 0.3 ppm with peaks at 0.6 ppm. The patterns were not consistent.	**Confirmed findings.** Models in hyposensitives and hypersensitives yielded BMDs of 1.66 ppm and 0.71 ppm, respectively, higher than the maximum concentration suggesting caution in using such result for POD derivation. BMDLs were 0.98 ppm and 0.68 ppm, in hyposensitives and hypersensitives respectively.
SPES sum score	Statistically significant increase in hypersensitives at 0.3/0.6 ppm ($p < 0.001$) and 0.4/0.8 ppm ($p < 0.01$) compared to pre-exposure.	**Confirmed findings.** Models in hyposensitives yielded higher BMD and BMDL than in hypersensitives, and higher than the maximum concentration suggesting caution in using such result for POD derivation. BMDLs were 1.02 and 0.71 ppm, in hyposensitives and hypersensitives respectively.
SPES eye irritation	Mean scores generally increased after exposure. No statistically significant changes.	**Confirmed** the reported lack of significant findings, except for a significant increase in hypersensitives exposed to 0.3 ppm with 0.6 ppm peaks, and a borderline significant increase (p-value =0.05) in hypersensitives exposed to 0.5 ppm.

Endpoint	Published result	ICF Result
		No model was viable for the data in hyposensitive volunteers. The estimated BMD and BMDLs in hypersensitive volunteers are 8.38 ppm and 0.75 ppm, respectively, higher than the maximum concentration suggesting caution in using such result for POD derivation.
SPES nasal irritation	Mean scores both increased and decreased after exposure for different concentrations and sensitivity group combinations. No statistically significant changes.	**Confirmed** the reported lack of significant findings. BMD (BMDLs) were 2.56 (1.10) ppm and 0.94 (0.77) ppm, in hyposensitives and hypersensitives respectively. All are higher than the maximum concentration suggesting caution in using such result for POD derivation.
SPES olfactory symptoms	Increased symptoms compared to the control concentration were found at each of the concentration; statistically significant for hypersensitive volunteers exposed at 0.4 ppm with peaks at 0.8 ppm. A concentration-effect relationship was not found. Hypersensitive volunteers reported consistently higher complaints than hyposensitive volunteers, statistically significantly higher at 0.3 ppm with 0.6 ppm peaks and at 0.4 ppm with 0.8 ppm peaks.	**Confirmed** statistically significant differences observed at all concentrations above the control condition for the hypersensitive group, and at 0.4 ppm with 0.8 ppm peaks and 0.5 ppm in the hyposensitive group. **Confirmed** the significant findings when comparing hypersensitive volunteers with hyposensitive volunteers at concentrations with peaks. When comparing to control condition, no significant changes were observed in our analyses. No model was viable for the data in hyposensitive volunteers. The estimated BMD and BMDLs in hypersensitive volunteers are 0.76 ppm and 0.71 ppm, respectively, higher than the maximum concentration suggesting caution in using such result for POD derivation.
SPES perception of impure air	Increase in hypersensitive at all exposure levels (including clean air, 0.01 ppm); increase at doses of 0.3 ppm with 0.6 ppm peaks and 0.5 ppm for hyposensitives	**Confirmed** the reported significant changes between post- and pre-exposure, and between hypo- and hypersensitive volunteers. No model was viable for the data in hyposensitive volunteers. The estimated BMD and BMDLs in hypersensitive volunteers are 2.42 ppm and 0.89 ppm, respectively, higher than the maximum

Endpoint	Published result	ICF Result
		concentration suggesting caution in using such result for POD derivation.
Endpoints without ICF Analyses		
Conjunctival redness	Statistically significantly increased redness after exposure for hyposensitives at 0 ppm, and statistically significantly decreased redness after exposure for hypersensitives at 0 ppm. No significant differences in exposed vs. control.	Unable to verify. It is not clear how this analysis was done, since the tabulated numbers seem to be the percentages of subjects where the redness either decreased, remained constant, or increased, rather than the percentage changes in the average redness. It is also unclear why the same percentage decrease of 23.81% for hypersensitives is significant at a dose of 0 ppm but not at 0.4 ppm with 4 peaks at 0.8 ppm. In the absence of an explanation for the data in Table Online Resource 3, it is not feasible to check this analysis.

3. Methods

Analyses using summary measures need several caveats. First, one must assume independence between responses at different doses. In both studies evaluated this assumption is violated by the controlled design. Second, use of digitized data from figures such as box plots or line bars (as was the case for Lang et al., (2008)) is likely to introduce errors. The analyses described below used several statistical approaches. Rates were compared using Fisher's exact test and the Cochran-Armitage trend test was used to evaluate trends in rates of conjunctival redness with and without ethyl acetate (EA). Mean differences before and after exposure, or between any exposure scenario and control condition were tested using Student t-test. When data was provided in box plots, standard deviations were calculated assuming that the data are reasonably represented by a normal distribution, the width of the IQR is approximately 1.35 SDs (Rosner, 2015). Thus, SD can be calculated as IQR/1.35.

The Spearman rank correlation coefficient and corresponding p-values were calculated to evaluate possible differences between blinking frequency and the subjective rating of ocular irritation from Lang et al. (2008). For all statistical tests $p < 0.05$ was considered statistically significant. Adjustments for multiple comparisons were not performed.

Some of the endpoints were modelled using EPA's Benchmark Dose Software (BMDS, Version 3.3rc10, https://www.epa.gov/bmds) to fit and plot dose-response models, and to estimate the BMD as the dose at which there was a 10% extra risk above an assumed 0% risk for unexposed subjects. The BMD approach is to fit statistical models for the probability of a response as a function of the dose. A variety of statistical models are fitted to the data and the best-fitting statistical model is selected as the one with the lowest Akaike Information Criterion (AIC) statistic We also estimated the BMDL, defined as a one-sided 95% lower confidence limit for the BMD.

Analyses were performed using R software version 3.4.4 (R Core Team, 2013), SAS software version 9.4 (SAS,2013) and BMDS Version 3.3rc10 (EPA, 2022b). Appendix 1 includes the Excel file with all the

datasets used. SAS code and listing file used for the Excel input file, Fisher exact tests, Cochran-Armitage tests, and sign test reported in this memorandum. The Excel output files containing the detailed BMDS Version 3.3rc10 dose-response analyses summarized in this memorandum are also attached in Appendix 4. Data from figures were digitized using the WebPlotDigitizer tool (https://apps.automeris.io/wpd/).

4. Results

Lang et al. (2008)

Conjunctival Redness

Lang et al. (2008) reported that conjunctival redness significantly increased at 0.5 ppm with peaks of 1.0 ppm. ICF digitized the data illustrated in **Figure 1** and compared the percentage of subjects showing moderate (grade 3) redness of the eyes to the control condition at 15, 120 and 195 minutes after start of exposure. The Fisher's exact confirmed the reported finding that eye redness rates were significantly different across the different concentrations only after 195 minutes after exposure (p-value =0.048). There was no difference in eye redness rates at 15 or 120 minutes after exposure (p-values of 0.367, and 0.216, respectively).

The Cochran-Armitage trend test indicated that there was no evidence of an increasing trend in eye redness rates across doses at 15, 120 and 195 minutes after start of exposure (p-values of 0.238, 0.238 and 0.096, respectively).

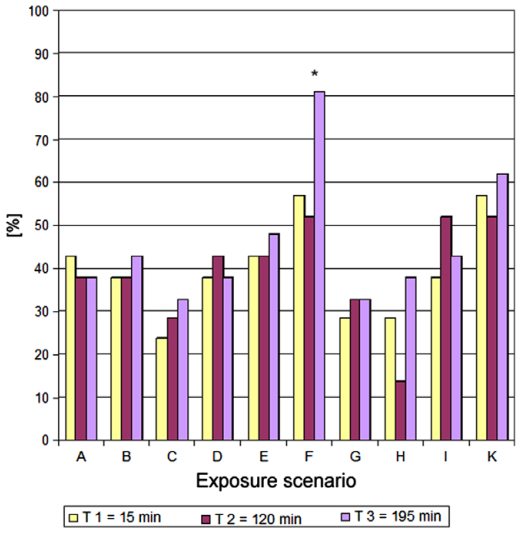

Figure 1. Conjunctival redness in 21 subjects during exposure to different concentrations of formaldehyde with or without ethyl acetate (EA). Results are expressed as percentage of subjects showing moderate (grade 3) redness of the eyes. A = 0 ppm, B = 0.15 ppm, C = 0.3 ppm, D = 0.3 ppm + 4 peaks of 0.6 ppm, E = 0.5 ppm, F = 0.5 ppm + 4 peaks of 1.0 ppm, G = 0 ppm + EA, H = 0.3 ppm + EA, I = 0.5 ppm + EA, K = 0.5 ppm + 4 peaks of 1.0 ppm + EA. Statistics: McNamar's test of symmetry; *p < .05. Figure 2 from Lang et al. (2008).

Benchmark Dose-response Modeling, with and without EA

Table 2 presents the BMD modelling results of the eye redness experiments at 195 minutes after exposure. The selected model based on the AIC are the Weibull and the multistage degree 2, for concentrations without and with EA, respectively. The selected BMDs (BMDLs) are 0.49 (0.18) ppm at 15 minutes after exposure for concentrations without EA, and 0.23 (0.08) ppm at 120 minutes after exposure for doses with EA.

Table 2. Summary of benchmark dose modeling results for conjunctival redness.

Time after start of exposure	Model[a]	BMD (ppm)	BMDL (ppm)	Notes

T1 - 15 min	**Weibull**	**0.494191**	**0.182096**	**Lowest AIC**
T2 - 120 min	Weibull	0.494188	0.140721	Lowest AIC; BMD/BMDL ratio > 3
T3 - 195 min	Weibull	0.48736	0.110082	Lowest AIC; BMD/BMDL ratio > 3
T1 - 15 min; Dose + EA	Multistage Degree 1	0.226009	0.069051	Lowest AIC; BMD/BMDL ratio > 3
T2 - 120 min; Dose + EA	**Multistage Degree 2**	**0.229524**	**0.080274**	**Lowest AIC**
T3 - 195 min; Dose + EA	Multistage Degree 2	0.22202	0.059534	Lowest AIC; BMD/BMDL ratio > 3

AIC = Akaike information criterion; BMD = benchmark dose; BMDL = benchmark dose lower limit.
[a] Selected model in bold.

Blinking Frequency

Lang et al. (2008) reported a statistically significant increase in blinking frequency was observed 195 min after the start of exposure in subjects exposed to 0.5 ppm formaldehyde with peaks of 1.0 ppm with or without EA compared to the 0 ppm-control condition with or without EA. In absence of raw data, ICF compared the mean blinking frequencies between any exposure scenario and the 0 ppm-control condition with or without EA were using Student t-test. No statistically significant differences were observed (**Table 3**).

Table 3. Blinking frequency per 90s in subjects during exposure (t = 195 min) to different concentrations of formaldehyde with or without EA.

			p-values (compared to 0 ppm)		p-values (compared to 0 ppm + EA)	
FA, ppm	**EA**	**Mean (SD)**	*Published*	*ICF*	*Published*	*ICF*
0	No	28.2 (30.2)	–	–	NR	0.97
	Yes	28.6 (30.9)	NR	0.97	–	–
0.15	No	31.2 (31.4)	NR	0.75	NR	0.79
0.3	No	27.8 (24.7)	NR	0.96	NR	0.93
	Yes	29.6 (24.0)	NR	0.87	NR	0.91
0.3/0.6[a]	No	34.4 (23.6)	NR	0.46	NR	0.50
0.5	No	29.2 (29.7)	NR	0.91	NR	0.95
	Yes	34.5 (35.1)	NR	0.54	NR	0.57
0.5/1.0	No	46.3 (45.6)	<0.05	0.14	<0.05	0.15
	Yes	45.2 (45.0)	<0.05	0.16	<0.05	0.17

FA = formaldehyde; EA = ethyl acetate; SD = standard deviation; NR = not reported
[a] Numbers given after a forward slash indicate concentrations for four 15-minute peaks.

Benchmark Dose-response Modeling, with and without EA

Table 4 presents the BMD modelling results of the blinking frequency at concentrations without peaks, with and without EA. The results using summary data yield BMDs and BMDLs higher than the maximum dose, caution in using the results for POD derivation.

Table 4. Benchmark dose modeling for blinking frequency.

Exposure Group	Model	BMD (ppm)	BMDL (ppm)	Notes
FA	Polynomial Degree 3	4.932258	0.653476	Lowest AIC; BMD/BMDL ratio > 3; BMD higher than maximum dose; BMDL higher than maximum dose
FA + EA	Polynomial Degree 2	1.101388	0.594075	Lowest AIC; BMD higher than maximum dose; BMDL higher than maximum dose

FA = formaldehyde; EA = ethyl acetate; BMD = benchmark dose; BMDL = benchmark dose level; AIC = Akaike information criterion.

Reaction Times

Visual Stimulus

Lang et al. (2008) reported a significantly increased decision reaction time upon a visual stimulus in subjects exposed to 0.3 ppm with or without co-exposure to EA. ICF digitized the data illustrated in **Figure 2** and compared mean reaction times using Student's test. No statistically significant differences were observed (**Table 5**).

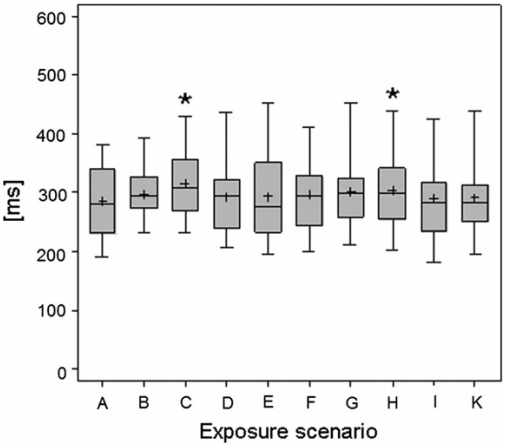

Figure 2. Decision reaction time upon a visual stimulus after exposure to different concentrations of formaldehyde with or without ethyl acetate (EA). Results are expressed in a box-whisker plot. A = 0 ppm, B = 0.15 ppm, C = 0.3 ppm, D = 0.3 ppm + 4 peaks of 0.6 ppm, E = 0.5 ppm, F = 0.5 ppm + 4 peaks of 1.0 ppm, G = 0 ppm + EA, H = 0.3 ppm + EA, I = 0.5 ppm + EA, K = 0.5 ppm + 4 peaks of 1.0 ppm + EA. Statistics: repeated measures ANOVA with contrasts; *p < 0.05. Figure 3 from Lang et al. (2008).

Table 5. Summary of results for reaction time to visual stimulus.

			p-values compared to 0 ppm	
FA, ppm	**EA**	**Mean (SD)**	*Published*	*ICF*
0	No	285 (80)	–	–
	Yes	301 (50)	NR	0.45
0.15	No	296 (40)	NR	0.57
0.3	No	315 (64)	<0.05	0.19
	Yes	304 (64)	<0.05	0.40
0.3/0.6[a]	No	294 (61)	NR	0.69
0.5	No	294 (89)	NR	0.73
	Yes	290 (61)	NR	0.82
0.5/1.0	No	295 (63)	NR	0.66
	Yes	292 (45)	NR	0.74

FA = formaldehyde; EA = ethyl acetate; SD = standard deviation; NR = not reported

[a] Numbers given after a forward slash indicate concentrations four 15-minute peaks.

Benchmark Dose-response Modeling, with and without EA

Table 6 presents the BMD modelling results of the visual stimulus at concentrations without peaks, with and without EA. No models were viable for concentrations without EA; for doses with EA, the results yield a BMD of 0.54 ppm and BMDL of 0.51 ppm, both slightly higher than the maximum dose, suggesting caution in using the results for POD derivation.

Table 6. Benchmark dose modeling for reaction time to visual stimulus.

Exposure Group	Model	BMD (ppm)	BMDL (ppm)	Notes
FA	No viable models	–	–	
FA + EA	Power	0.544835	0.509944	Lowest AIC; BMD higher than maximum dose, BMDL higher than maximum dose

FA = formaldehyde; EA = ethyl acetate; BMD = benchmark dose; BMDL = benchmark dose level; AIC = Akaike information criterion.

Auditory Stimulus

Lang et al. (2008) reported a significantly increased decision reaction time upon an acoustic stimulus in subjects exposed to 0.3 ppm without co-exposure to EA. ICF digitized the data illustrated in **Figure 3** and compared mean reaction times using Student's test. No statistically significant differences were observed (**Table 7**).

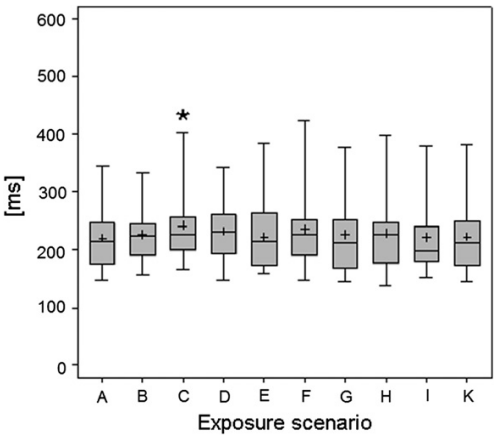

Figure 3. Decision reaction time upon an acoustic stimulus after exposure to different concentrations of formaldehyde with or without ethyl acetate (EA). Results are expressed in a box-whisker plot. A = 0 ppm, B = 0.15 ppm, C = 0.3 ppm, D = 0.3 ppm + 4 peaks of 0.6 ppm, E = 0.5 ppm, F = 0.5 ppm + 4 peaks of 1.0 ppm, G = 0 ppm + EA, H = 0.3 ppm + EA, I = 0.5 ppm + EA, K = 0.5 ppm + 4 peaks of 1.0 ppm + EA. Statistics: repeated measures ANOVA with contrasts; *p < 0.05. Figure 4 from Lang et al. (2008).

Table 7. Summary of results for reaction time to auditory stimulus.

			p-values compared to 0 ppm	
FA, ppm	**EA**	**Mean (SD)**	*Published*	*ICF*
0	No	218 (54)	–	–
	Yes	225 (62)	NR	0.67
0.15	No	224 (40)	NR	0.65
0.3	No	239 (42)	<0.05	0.16
	Yes	224 (52)	NR	0.69
0.3/0.6[a]	No	230 (50)	NR	0.45
0.5	No	220 (67)	NR	0.91
	Yes	220 (45)	NR	0.89
0.5/1.0	No	234 (45)	NR	0.28
	Yes	220 (56)	NR	0.90

FA = formaldehyde; EA = ethyl acetate; SD = standard deviation; NR = not reported

[a] Numbers given after a forward slash indicate concentrations for four 15-minute peaks.

Benchmark Dose-response Modeling, with and without EA

Table 8 presents the BMD modelling results of the auditory stimulus at concentrations without peaks, with and without EA. The results using concentrations without EA data yield BMDs and BMDLs higher than the maximum dose, suggesting caution in using the results for POD derivation. For doses with EA co-exposure, the estimated BMD is 1.5 ppm with a BMDL of 0.64, both higher than the maximum administered dose.

Table 8. Benchmark dose modeling for reaction time to auditory stimulus.

Exposure Group	Model	BMD (ppm)	BMDL (ppm)	Notes
FA	Polynomial Degree 2	5.284559	0.757491	Lowest AIC; BMD/BMDL ratio > 3; BMD higher than maximum dose; BMDL higher than maximum dose
FA + EA	Polynomial Degree 3	1.506023	0.63936	Lowest AIC; BMD higher than maximum dose; BMDL higher than maximum dose

FA = formaldehyde; EA = ethyl acetate; BMD = benchmark dose; BMDL = benchmark dose level; AIC = Akaike information criterion.

Combined Visual/Auditory Stimulus

Lang et al. (2008) reported a significantly increased combined visual and acoustic reaction time at 0.3 ppm without EA. ICF digitized the data illustrated in **Figure 4** and compared mean reaction times using Student's test. No statistically significant differences were observed (**Table 9**).

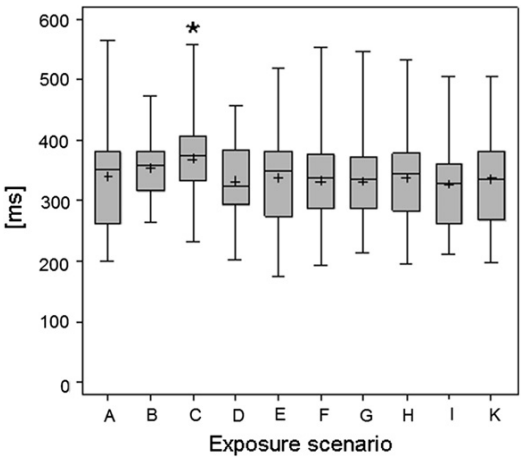

Figure 4. Decision reaction time upon a combined visual/auditory stimulus after exposure to different concentrations of formaldehyde with or without ethyl acetate (EA). Results are expressed in a box-whisker plot. A = 0 ppm, B = 0.15 ppm, C = 0.3 ppm, D = 0.3 ppm + 4 peaks of 0.6 ppm, E = 0.5 ppm, F = 0.5 ppm + 4 peaks of 1.0 ppm, G = 0 ppm + EA, H = 0.3 ppm + EA, I = 0.5 ppm + EA, K = 0.5 ppm + 4 peaks of 1.0 ppm + EA. Statistics: repeated measures ANOVA with contrasts; *p < 0.05.Figure 5 from Lang et al. (2008).

Table 9. Summary of results for reaction time to combined stimulus.

			p-values compared to 0 ppm	
FA, ppm	**EA**	**Mean (SD)**	*Published*	*ICF*
0	No	339 (88)	–	–
	Yes	331 (63)	NR	0.71
0.15	No	354 (48)	NR	0.52
0.3	No	367 (54)	<0.05	0.23
	Yes	337 (71)	NR	0.93
0.3/0.6[a]	No	331 (67)	NR	0.72
0.5	No	337 (80)	NR	0.93
	Yes	326 (73)	NR	0.60
0.5/1.0	No	331 (67)	NR	0.72
	Yes	337 (84)	NR	0.93

FA = formaldehyde; EA = ethyl acetate; SD = standard deviation; NR = not reported

[a] Numbers given after a forward slash indicate concentrations with four 15-minute peaks.

Benchmark Dose-response Modeling, with and without EA

Table 10 presents the BMD modelling results of the combined stimulus at concentrations without peaks, with and without EA. No model was viable for modelling of responses at concentrations without EA co-exposure. For doses with EA co-exposure, the estimated BMD is 0.56 ppm with a BMDL of 0.51 ppm, both slightly higher than the maximum administered dose.

Table 10. Benchmark dose modeling for reaction time to combined stimulus.

Exposure Group	Model	BMD	BMDL	Notes
FA	No viable models	–	–	
FA + EA	Power	0.564214	0.515127	Lowest AIC; BMD higher than maximum dose; BMDL higher than maximum dose

FA = formaldehyde; EA = ethyl acetate; BMD = benchmark dose; BMDL = benchmark dose level; AIC = Akaike information criterion.

Subjective Symptoms

Total Symptom Score

Lang et al. (2008) reported that the mean total symptom score was significantly higher versus the controls at 0.5 ppm with peaks, with or without EA. ICF digitized the data illustrated in **Figure 5** and compared mean total symptom score using Student's test (**Table 11**). Compared to the 0 ppm-control group, significant increases in total symptom scores were observed at all concentrations except at 0.15 ppm and 0.3 ppm without EA co-exposure. Compared to the 0 ppm with EA co-exposure, mean total symptom score was significantly higher only at concentrations of 0 ppm and 0.15 ppm FA without EA co-exposure. However, the results need to be interpreted with extreme caution give the clear violation of normality illustrated in Figure 5.

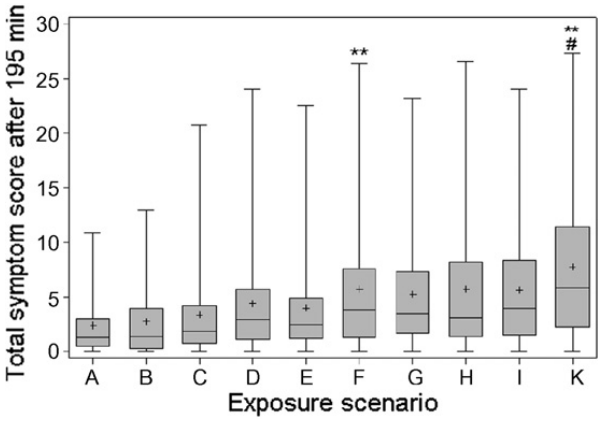

Figure 5. Total symptom score recorded during exposure (t = 195 min) to different concentrations of formaldehyde with or without ethyl acetate (EA). Results are expressed in a box-whisker plot. A = 0 ppm, B = 0.15 ppm, C = 0.3 ppm, D = 0.3 ppm + 4 peaks of 0.6 ppm, E = 0.5 ppm, F = 0.5 ppm + 4 peaks of 1.0 ppm, G = 0 ppm + EA, H = 0.3 ppm + EA, I = 0.5 ppm + EA, K = 0.5 ppm + 4 peaks of 1.0 ppm + EA. Statistics: repeated measures ANOVA with contrasts; **p < 0.01 compared to 0 ppm; #p < 0.05 compared to 0 ppm + EA. Figure 5 from Lang et al. (2008).

Table 11. Summary of results for total symptoms during exposure (t = 195 min).

			p-values compared to 0 ppm		p-values compared to 0 ppm + EA	
FA, ppm	**EA**	**Mean (SD[a])**	*Published*	*ICF*	*Published*	*ICF*
0	No	2.386 (1.9)	–	–	NR	<0.01
	Yes	5.170 (4.2)	NR	<0.01	–	–
0.15	No	2.727 (2.7)	NR	0.68	NR	0.03
0.3	No	3.352 (2.5)	NR	0.15	NR	0.10
	Yes	5.682 (5.1)	NR	<0.01	NR	0.73
0.3/0.6[b]	No	4.432 (3.4)	NR	0.02	NR	0.50
0.5	No	3.920 (2.7)	NR	<0.05	NR	0.24
	Yes	5.625 (5.1)	NR	<0.05	NR	0.78
0.5/1.0	No	5.682 (4.6)	<0.01	<0.01	NR	0.71
	Yes	7.727 (6.8)	<0.01	<0.01	<0.05	0.16

FA = formaldehyde; EA = ethyl acetate; SD = standard deviation; NR = not reported

[a] Standard deviation calculated as IQR/1.35. Normality assumption clearly violated based on Figure 5.

[b] Numbers given after a forward slash indicate concentrations with four 15-minute peaks.

Benchmark Dose-response Modeling, with and without EA

Table 12 presents the BMD modelling results of the total symptom score at concentrations without peaks, with and without EA. The estimated BMDs are higher than the maximum concentration suggesting caution in using such result for POD derivation.

Table 12. Benchmark dose modeling for total symptoms.

Exposure Group	Model	BMD (ppm)	BMDL (ppm)	Notes
FA	Linear	0.763989	0.436026	Lowest AIC; BMD higher than maximum dose
FA + EA	Polynomial Degree 2	4.80752	0.655691	Lowest AIC; BMD/BMDL ratio > 3; BMD higher than maximum dose; BMDL higher than maximum dose

FA = formaldehyde; EA = ethyl acetate; BMD = benchmark dose; BMDL = benchmark dose level; AIC = Akaike information criterion.

Eye Irritation

Lang et al. (2008) reported that the mean eye irritation symptom score was significantly higher versus the controls in most cases. ICF digitized the data illustrated in **Figure 6** and compared mean total symptom score using Student's test (**Table 13**). Compared to the 0 ppm-control group, significant increases in eye irritation scores were observed at concentrations of 0.3 ppm and 0.5 ppm with peaks without EA co-exposure, and at 0.5 ppm with or without peaks, with EA co-exposure. Compared to the 0 ppm + EA co-exposure, mean eye irritation score was significantly higher only at concentrations of 0.5 ppm with four 1.0 ppm peaks with or without EA co-exposure.

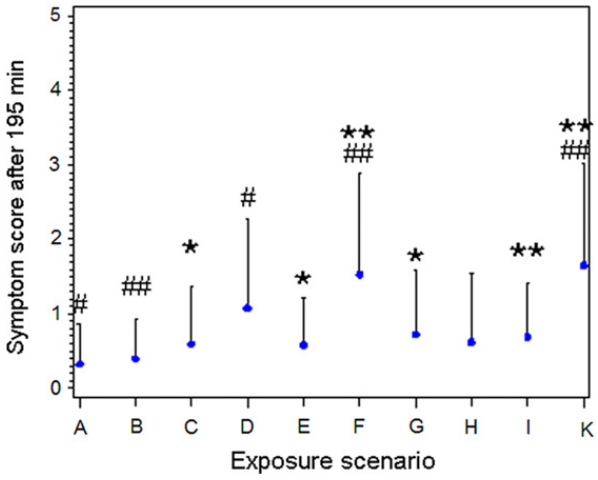

Figure 6. Symptom score for eye irritation recorded during exposure (t = 195 min) to different concentrations of formaldehyde with or without ethyl acetate (EA). Results are expressed in means ± SD. A = 0 ppm, B = 0.15 ppm, C = 0.3 ppm, D = 0.3 ppm + 4 peaks of 0.6 ppm, E = 0.5 ppm, F = 0.5 ppm + 4 peaks of 1.0 ppm, G = 0 ppm + EA, H = 0.3 ppm + EA, I = 0.5 ppm + EA, K = 0.5 ppm + 4 peaks of 1.0 ppm + EA. Statistics: repeated measures ANOVA with contrasts; *p < 0.05 and **p < 0.01 compared to 0 ppm; #p < 0.05 and ##p < 0.01 compared to 0 ppm + EA. Figure 7 from Lang et al. (2008).

Table 13. Summary of results for eye irritation recorded during exposure (t = 195 min).

			p-values compared to 0 ppm		p-values compared to 0 ppm + EA	
FA, ppm	**EA**	**Mean (SD)**	*Published*	*ICF*	*Published*	*ICF*
0	No	0.321 (0.5)	–	–	<0.05	0.08
	Yes	0.714 (0.9)	<0.05	0.08	–	–
0.15	No	0.383 (0.5)	NR	0.52	<0.01	0.19
0.3	No	0.582 (0.8)	<0.05	0.14	NR	0.71
	Yes	0.602 (0.9)	NR	0.19	NR	0.72
0.3/0.6[a]	No	1.066 (1.2)	NR	<0.01	<0.05	0.23
0.5	No	0.571 (0.6)	<0.05	0.09	NR	0.67
	Yes	0.684 (0.7)	<0.01	<0.05	NR	1.00

FA, ppm	EA	Mean (SD)	p-values compared to 0 ppm		p-values compared to 0 ppm + EA	
			Published	*ICF*	*Published*	*ICF*
0.5/1.0	No	1.52 (1.4)	<0.01	<0.001	<0.01	<0.05
	Yes	1.648 (1.4)	<0.01	<0.001	<0.01	<0.05

FA = formaldehyde; EA = ethyl acetate; SD = standard deviation; NR = not reported

[a] Numbers given after a forward slash indicate concentrations with four 15-minute peaks.

Benchmark Dose-response Modeling, with and without EA

Table 14 presents the BMD modelling results of the eye irritation symptom score at concentrations without peaks, with and without EA. The estimated BMDs and BMDLs are higher than the maximum concentration suggesting caution in using such result for POD derivation.

Table 14. Benchmark dose modeling for eye irritation.

Exposure Group	Model	BMD (ppm)	BMDL (ppm)	Notes
FA	Polynomial Degree 2	1.098926	0.52991	Lowest AIC; BMD higher than maximum dose; BMDL higher than maximum dose
FA + EA	Power	9.666709	0.52028	Lowest AIC; BMD/BMDL ratio > 3; BMD higher than maximum dose; BMDL higher than maximum dose

FA = formaldehyde; EA = ethyl acetate; BMD = benchmark dose; BMDL = benchmark dose level; AIC = Akaike information criterion.

Nasal Irritation

Lang et al. (2008) observed significantly higher mean symptom scores for nasal irritation at 0.5 ppm with peaks of 1.0 ppm with or without EA compared to either control, and at 0.3 ppm + EA or 0.5 ppm + EA compared to 0 ppm-control. ICF digitized the data illustrated in **Figure 7** and compared mean symptom scores using Student's test (**Table 15**). Our analyses confirmed the finding of significantly higher mean symptom scores for nasal irritation at 0.5 ppm with peaks of 1.0 ppm with or without EA compared to either control, and at 0.3 ppm + EA or 0.5 ppm + EA compared to 0 ppm-control. In addition, a significantly higher mean score was observed at 0.15 ppm without EA compared to 0 ppm + EA.

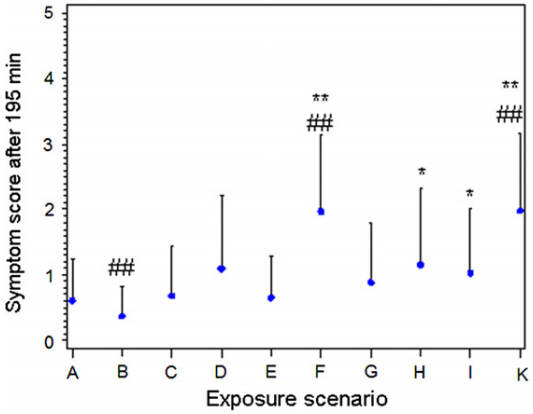

Figure 7. Symptom score for nasal irritation recorded during exposure (t = 195 min) to different concentrations of formaldehyde with or without ethyl acetate (EA). Results are expressed in means ± SD. A = 0 ppm, B = 0.15 ppm, C = 0.3 ppm, D = 0.3 ppm + 4 peaks of 0.6 ppm, E = 0.5 ppm, F = 0.5 ppm + 4 peaks of 1.0 ppm, G = 0 ppm + EA, H = 0.3 ppm + EA, I = 0.5 ppm + EA, K = 0.5 ppm + 4 peaks of 1.0 ppm + EA. Statistics: repeated measures ANOVA with contrasts; *p < 0.05 and **p < 0.01 compared to 0 ppm; ##p < 0.01 compared to 0 ppm + EA. Figure 8 from Lang et al. (2008).

Table 15. Summary of results for nasal irritation scores recorded during exposure (t = 195 min).

FA, ppm	EA	Mean (SD[a])	p-values compared to 0 ppm		p-values compared to 0 ppm + EA	
			Published	*ICF*	*Published*	*ICF*
0	No	0.599 (0.6)	–	–	NR	0.21
	Yes	0.885 (0.9)	NR	0.21	–	–
0.15	No	0.371 (0.4)	NR	0.21	<0.01	<0.05

FA, ppm	EA	Mean (SD[a])	p-values compared to 0 ppm		p-values compared to 0 ppm + EA	
			Published	*ICF*	*Published*	*ICF*
0.3	No	0.676 (0.8)	NR	0.65	NR	0.45
	Yes	1.151 (1.2)	<0.05	<0.05	NR	0.36
0.3/0.6[a]	No	1.094 (1.1)	NR	0.07	NR	0.52
0.5	No	0.657 (0.6)	NR	0.59	NR	0.40
	Yes	1.028 (1.0)	<0.05	0.12	NR	0.74
0.5/1.0	No	1.970 (1.2)	<0.01	<0.001	<0.01	<0.01
	Yes	1.989 (1.2)	<0.01	<0.001	<0.01	<0.01

FA = formaldehyde; EA = ethyl acetate; SD = standard deviation; NR = not reported

[a] Numbers given after a forward slash indicate concentrations with four 15-minute peaks.

Benchmark Dose-response Modeling, with and without EA

Table 16 presents the BMD modelling results of the nasal irritation ratings at concentrations without peaks, with and without EA. The estimated BMDs and BMDLs are higher than the maximum concentration suggesting caution in using such result for POD derivation.

Table 16. Benchmark dose modeling for nasal irritation.

Exposure Group	Model	BMD (ppm)	BMDL (ppm)	Notes
FA	Polynomial Degree 2	0.98341	0.589824	Lowest AIC; BMD higher than maximum dose; BMDL higher than maximum dose
FA + EA	Power	3.029298	0.523519	Lowest AIC; BMD/BMDL ratio > 3; BMD higher than maximum dose; BMDL higher than maximum dose

FA = formaldehyde; EA = ethyl acetate; BMD = benchmark dose; BMDL = benchmark dose level; AIC = Akaike information criterion.

Olfactory Symptoms

Lang et al. (2008) reported that significantly increased scores of olfactory symptoms were observed at concentrations of 0.3 ppm FA and higher, and peak exposure and co-exposure with EA increased scores. In addition, only at 0.5 ppm with peaks and EA scores were significantly higher than the 0 ppm +EA control. ICF digitized the data illustrated in **Figure 8** and compared mean symptom scores using Student's test (**Table 17**). Our analyses confirmed the reported findings, except for a lack of significant differences at 0.5 ppm compared to 0 ppm, and at 0.5 ppm with peaks and EA compared to 0 ppm +EA.

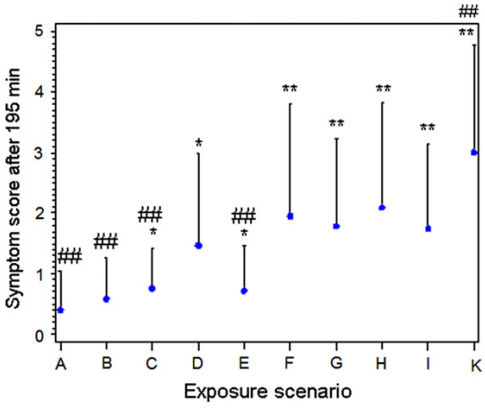

Figure 8. Olfactory symptom score recorded during exposure (t = 195 min) to different concentrations of formaldehyde with or without ethyl acetate (EA). Results are expressed in means ± SD. A = 0 ppm, B = 0.15 ppm, C = 0.3 ppm, D = 0.3 ppm + 4 peaks of 0.6 ppm, E = 0.5 ppm, F = 0.5 ppm + 4 peaks of 1.0 ppm, G = 0 ppm + EA, H = 0.3 ppm + EA, I = 0.5 ppm + EA, K = 0.5 ppm + 4 peaks of 1.0 ppm + EA. Statistics: repeated measures ANOVA with contrasts; *p < 0.05 and **p < 0.01 compared to 0 ppm; ##p < 0.01 compared to 0 ppm + EA. Figure 9 from Lang et al. (2008).

Table 17. Summary of results for olfactory symptoms scores recorded during exposure (t=195 min).

			p-values compared to 0 ppm		p-values compared to 0 ppm + EA	
FA, ppm	**EA**	**Mean (SD)**	*Published*	*ICF*	*Published*	*ICF*
0	No	0.402 (0.6)	–	–	<0.01	<0.001
	Yes	1.783 (1.4)	<0.01	<0.001	–	–
0.15	No	0.578 (0.7)	NR	0.33	<0.01	<0.01
0.3	No	0.754 (0.7)	<0.05	0.05	<0.01	<0.01
	Yes	2.086 (1.7)	<0.01	<0.001	NR	0.54
0.3/0.6[a]	No	1.459 (1.5)	<0.05	<0.01	NR	0.51
0.5	No	0.725 (0.7)	<0.05	0.14	<0.01	<0.01
	Yes	1.753 (1.4)	<0.01	<0.001	NR	1.00
0.5/1.0	No	1.939 (1.9)	<0.01	<0.01	NR	0.85
	Yes	2.997 (1.8)	<0.01	<0.001	<0.01	0.2

FA = formaldehyde; EA = ethyl acetate; SD = standard deviation; NR = not reported

[s] Numbers given after a forward slash indicate concentrations with four 15-minute peaks.

Benchmark Dose-response Modeling, with and without EA

Table 18 presents the BMD modelling results of the olfactory symptoms at concentrations without peaks, with and without EA. For concentrations without EA, the estimated BMD is 1.00 ppm with a BMDL of 0.51 ppm, both higher than the maximum concentration suggesting caution in using such result for POD derivation. No viable model converged at concentrations with EA co-exposure.

Table 18. Benchmark dose modeling for olfactory symptoms.

Exposure Group	Model	BMD (ppm)	BMDL (ppm)	Notes
FA	Polynomial Degree 2	1.003767	0.506402	Lowest AIC; BMD higher than maximum dose; BMDL higher than maximum dose
FA + EA	No viable models	–	–	

FA = formaldehyde; EA = ethyl acetate; BMD = benchmark dose; BMDL = benchmark dose level; AIC = Akaike information criterion.

Respiratory Symptoms

Lang et al. (2008) reported significantly increased scores of respiratory symptoms were observed at concentrations of 0.3 ppm FA without peaks, at 0.5 ppm with peaks, and at all levels with EA co-exposures. ICF digitized the data illustrated in **Figure 9** and compared mean symptom scores using Student's test (**Table 19**). Our analyses confirmed only a single finding, of significant increase in mean score at 0.5 ppm with peaks and EA co-exposure compared to the 0 ppm-control.

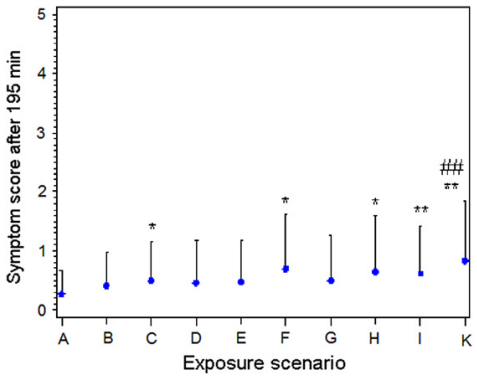

Figure 9. Respiratory symptom score recorded during exposure (t = 195 min) to different concentrations of formaldehyde with or without ethyl acetate (EA). Results are expressed in means ± SD. A = 0 ppm, B = 0.15 ppm, C = 0.3 ppm, D = 0.3 ppm + 4 peaks of 0.6 ppm, E = 0.5 ppm, F = 0.5 ppm + 4 peaks of 1.0 ppm, G = 0 ppm + EA, H = 0.3 ppm + EA, I = 0.5 ppm + EA, K = 0.5 ppm + 4 peaks of 1.0 ppm + EA. Statistics: repeated measures ANOVA with contrasts; *p < 0.05 and **p < 0.01 compared to 0 ppm; ##p < 0.01 compared to 0 ppm + EA. Figure 10 from Lang et al. (2008).

Table 19. Summary of results for respiratory symptoms recorded during exposure (t = 195 min).

			p-values compared to 0 ppm		p-values compared to 0 ppm + EA	
FA, ppm	**EA**	**Mean (SD)**	*Published*	*ICF*	*Published*	*ICF*
0	No	0.278 (0.4)	–	–	NR	0.31
	Yes	0.494 (0.8)	NR	0.31	–	–
0.15	No	0.404 (0.6)	NR	0.53	NR	0.65
0.3	No	0.494 (0.7)	<0.05	0.26	NR	1.00
	Yes	0.646 (1.0)	<0.05	0.21	NR	0.72
0.3/0.6[a]	No	0.449 (0.7)	NR	0.57	NR	0.67
0.5	No	0.467 (0.7)	NR	0.26	NR	1.00
	Yes	0.619 (0.8)	<0.01	0.13	NR	0.69
0.5/1.0	No	0.691 (0.9)	<0.05	0.07	NR	0.45
	Yes	0.826 (1.0)	<0.01	<0.05	<0.01	0.29

FA = formaldehyde; EA = ethyl acetate; SD = standard deviation; NR = not reported

[a] Numbers given after a forward slash indicate concentrations with four 15-minute peaks.

Benchmark Dose-response Modeling, with and without EA

Table 20 presents the BMD modelling results of the respiratory symptoms at concentrations without peaks, with and without EA. For concentrations without EA, the estimated BMD is 0.78 ppm (higher than the maximum concentration), with a BMDL of 0.35 ppm. For concentrations with EA co-exposure, the estimated BMD is 3.04 ppm with a BMDL of 0.64 ppm, both higher than the maximum concentration suggesting caution in using such result for POD derivation.

Table 20. Benchmark dose modeling for respiratory symptoms.

Exposure Group	Model	BMD (ppm)	BMDL (ppm)	Notes
FA	Power	0.777804	0.352226	Lowest AIC; BMD higher than maximum dose (non-constant variance)
FA + EA	Polynomial Degree 2	3.036719	0.635142	Lowest AIC; BMD/BMDL ratio > 3; BMD higher than maximum dose; BMDL higher than maximum dose

FA = formaldehyde; EA = ethyl acetate; BMD = benchmark dose; BMDL = benchmark dose level; AIC = Akaike information criterion.

Annoyance Ratings

Lang et al. (2009) reported that mean ratings of annoyance increased with elevated concentrations, and at all levels with co-exposure to EA, including the 0 ppm + EA control. ICF digitized the data illustrated in **Figure 10** and compared mean scores using Student's test (**Table 19**). Our analyses confirmed only a single finding, of significant increase in mean score at 0.3 ppm with EA co-exposure compared to the 0 ppm-control.

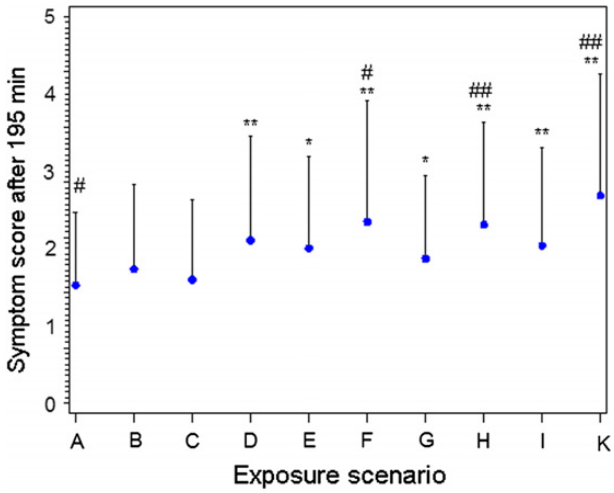

Figure 10. Annoyance ratings recorded during exposure (t = 195 min) to different concentrations of formaldehyde with or without ethyl acetate (EA). Results are expressed in means ± SD. A = 0 ppm, B = 0.15 ppm, C = 0.3 ppm, D = 0.3 ppm + 4 peaks of 0.6 ppm, E = 0.5 ppm, F = 0.5 ppm + 4 peaks of 1.0 ppm, G = 0 ppm + EA, H = 0.3 ppm + EA, I = 0.5 ppm + EA, K = 0.5 ppm + 4 peaks of 1.0 ppm + EA. Statistics: repeated measures ANOVA with contrasts; *p < 0.05 and **p < 0.01 compared to 0 ppm; #p < 0.05 and ##p < 0.01 compared to 0 ppm + EA. Figure 10 from Lang et al. (2008).

Table 21. Summary of results for annoyance ratings.

			p-values compared to 0 ppm		p-values compared to 0 ppm + EA	
FA, ppm	**EA**	**Mean (SD)**	*Published*	*ICF*	*Published*	*ICF*
0	No	1.526 (0.9)	–	–	<0.05	0.20
	Yes	1.873 (1.1)	<0.05	0.20	–	–
0.15	No	1.732 (1.1)	NR	0.52	NR	0.56
0.3	No	1.592 (1.0)	NR	0.74	NR	0.36
	Yes	2.313 (1.3)	<0.01	<0.05	<0.01	0.29
0.3/0.6[a]	No	2.107 (1.3)	<0.01	0.09	NR	0.59
0.5	No	2.004 (1.2)	<0.05	0.13	NR	0.78
	Yes	2.041 (1.3)	<0.01	0.16	NR	0.79
0.5/1.0	No	2.341 (1.6)	<0.01	0.05	<0.05	0.35
	Yes	2.678 (1.6)	<0.01	<0.01	<0.01	0.07

FA = formaldehyde; EA = ethyl acetate; SD = standard deviation; NR = not reported

[a] Numbers given after a forward slash indicate concentrations with four 15-minute peaks.

Benchmark Dose-response Modeling, with and without EA

Table 22 presents the BMD modelling results of the annoyance ratings symptoms at concentrations without peaks, with and without EA. For concentrations without EA, the estimated BMD is 0.52 ppm, with a BMDL of 0.51 ppm. For concentrations with EA co-exposure, the estimated BMD is 2.82 ppm with a BMDL of 0.73. The models yielded BMDs and BMDLs higher than the maximum concentration suggesting caution in using such result for POD derivation.

Table 22. Benchmark dose modeling for annoyance ratings.

Exposure Group	Model	BMD (ppm)	BMDL (ppm)	Notes
FA	Exponential 3	0.523935	0.511504	Lowest AIC; BMD higher than maximum dose; BMDL higher than maximum dose
FA + EA	Linear	2.822372	0.733003	Lowest AIC; BMD/BMDL ratio > 3; BMD higher than maximum dose; BMDL higher than maximum dose

FA = formaldehyde; EA = ethyl acetate; BMD = benchmark dose; BMDL = benchmark dose level; AIC = Akaike information criterion.

Correlations Between Subjective Eye Irritation and Blinking Frequency

Lang et al. (2008) reported a significant correlation between the subjectively rated eye irritation and the blinking frequency at the highest level tested (0.5 ppm with peaks) in the presence of EA (**Figure 11)**. The correlation was positive but not significant at 0.5 ppm with peaks of 1 ppm (with or without the one identified outlier) **Figure 12**. ICF digitized the data illustrated in **Figures 11** and **12** and calculated the Spearman rank correlations of 0.51 (p-value 0.02) and 0.36 (p-value =0.11). Our result confirmed the reported findings.

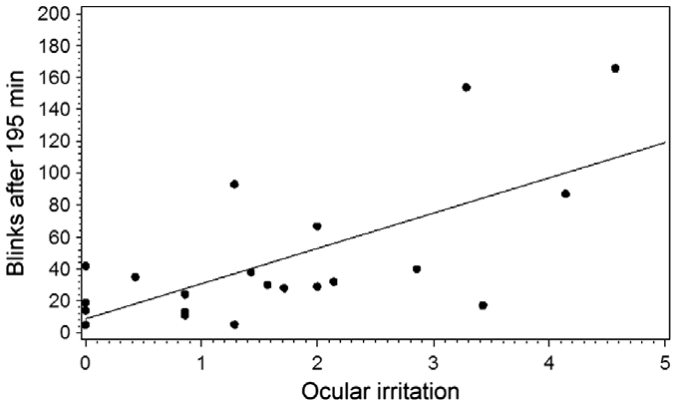

Figure 11. Correlation between subjectively rated eye irritation and the number of blinks in 90 s in subjects exposed to 0.5 ppm formaldehyde with peaks of 1.0 ppm and co-exposure to ethyl acetate (ρ = 0.54; y = 22.12x + 8.62; p = 0.01). After using the three sigma criteria, the two highest values were no outliers. A = 0 ppm, B = 0.15 ppm, C = 0.3 ppm, D = 0.3 ppm + 4 peaks of 0.6 ppm, E = 0.5 ppm, F = 0.5 ppm + 4 peaks of 1.0 ppm, G = 0 ppm + EA, H = 0.3 ppm + EA, I = 0.5 ppm + EA, K = 0.5 ppm + 4 peaks of 1.0 ppm + EA. Figure 12 from Lang et al. (2008).

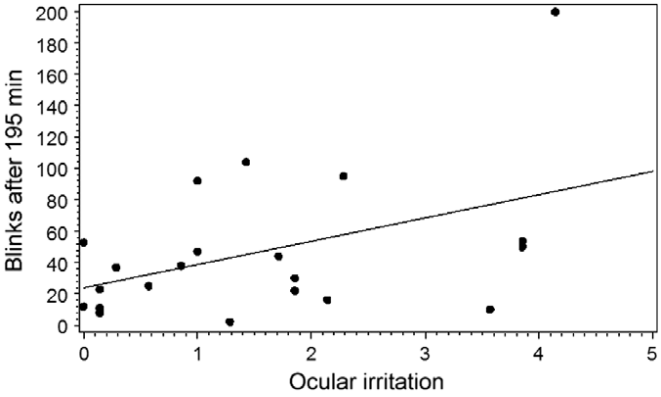

Figure 12. Correlation between subjectively rated eye irritation and the number of blinks in 90 s in subjects exposed to 0.5 ppm formaldehyde with peaks of 1.0 ppm (ρ = 0.36; y = 14.94x + 23.47; p = 0.10). The highest value in this graph is an outlier (after the three sigma criteria). Without this value, the correlation is ρ= 0.26 with a p-value of 0.26. Figure 13 from Lang et al. (2008).

Mueller et al. (2013)

Blinking Frequency

Mueller et al. (2013) reported a decrease in eye blinking frequency but only significant for hypersensitives exposed to 0 ppm. ICF used the data provided in the Online Resource 6 and compared mean total symptom score using Student's test (**Table 23**). Our analyses confirmed the previous findings. A consistent statistically significant change in eye-blinking frequency after exposure to formaldehyde was not observed. Although mean differences after exposure were generally negative reflecting a decrease in eye blinking frequency after exposure to FA, the change was only statistically significant among hypersensitives exposed to 0 ppm.

Table 23. Summary of results for eye-blinking frequency.

			p-values compared to pre-exposure	
FA, ppm	**Sensitivity Group**	**Mean Difference (SD)**	*Published*	*ICF*
0	Hypo	−2.30 (8.92)	NR	0.26
	Hyper	−4.05 (7.92)	<0.05	<0.05
0.3/0.6[a]	Hypo	−0.65 (5.96)	NR	0.63
	Hyper	−0.52 (7.26)	NR	0.74
0.4/0.8	Hypo	−1.95 (6.48)	NR	0.19
	Hyper	−1.52 (9.35)	NR	0.47
0.5	Hypo	−1.35 (4.49)	NR	0.19
	Hyper	−1.95 (8.19)	NR	0.28
0.7	Hypo	−1.90 (8.32)	NR	0.31
	Hyper	−3.86 (10.06)	NR	0.09

FA = formaldehyde; SD = standard deviation; NR = not reported

[a] Numbers given after a forward slash indicate concentrations with four 15-minute peaks.

Benchmark Dose-response Modeling

Table 24 presents the BMD modelling results of the mean change in eye blinking frequency at end of exposure compared to pre-exposure at concentrations without peaks, by sensitivity group. No model was viable for the data in hyposensitives. The estimated BMD and BMDLs in hypersensitives are 8.38 ppm and 0.72 ppm, respectively, higher than the maximum concentration suggesting caution in using such result for POD derivation.

Table 24. Benchmark dose modeling for eye-blinking frequency.

Sensitivity Group	Model	BMD (ppm)	BMDL (ppm)	Notes
Hypo	No viable models	–	–	
Hyper	Power	8.384901	0.722708	Lowest AIC; BMD/BMDL ratio > 3, BMD higher than maximum dose, BMDL higher than maximum dose

BMD = benchmark dose; BMDL = benchmark dose level; AIC = Akaike information criterion.

Tear Film Break-Up Time

Mueller at el. (2013) observed a significantly increased self-reported tear film break-up time at 0.4 ppm with peaks and 0.5 ppm ($p < 0.05$). Results of exposure to formaldehyde on changes in self-reported tear film break-up time are shown in in **Table 25** (using data from Table Online Resource 7 of Mueller et al. (2008)). Our analyses confirmed the findings of significant mean changes after exposure compared to pre-exposure at 0.4 ppm with peaks in both hyper- and hyposensitives, and at 0.5 ppm in hypersensitives. Statistically significantly decreases were also observed in hypersensitives when

compared to hyposensitives at 0 ppm. Our analyses did not observe the significant changes reported in Mueller et al. (2013) in hyposensitives at 0.3 ppm with peaks and at 0.7 ppm compared with the 0 ppm, but significant differences were observed hypersensitives at 0.4 ppm with peaks and at 0.5 ppm when compared to 0 ppm.

Table 25. Summary of results for tear film break-up time.

FA, ppm	Sensitivity Group	Mean Difference (SD)	p-values compared to pre-exposure		p-values compared to control condition		p-values compared to hyposensitives	
			Published	*ICF*	*Published*	*ICF*	*Published*	*ICF*
0	Hypo	54.67 (123.61)	NR	0.06	–	–	–	–
	Hyper	−3.71 (30.59)	NR	0.58	–	–	<0.01	<0.05
0.3/0.6[a]	Hypo	7.50 (30.87)	NR	0.28	<0.05	0.11	–	–
	Hyper	9.86 (39.45)	NR	0.26	NR	0.22	NR	0.83
0.4/0.8	Hypo	20.30 (42.91)	<0.05	<0.05	NR	0.25	–	–
	Hyper	26.63 (44.92)	<0.05	<0.01	NR	0.01	NR	0.65
0.5	Hypo	23.33 (64.18)	NR	0.11	NR	0.32	–	–
	Hyper	20.57 (43.28)	<0.05	<0.05	NR	<0.05	NR	0.87
0.7	Hypo	7.85 (52.29)	NR	0.51	<0.05	0.13	–	–
	Hyper	5.10 (45.70)	NR	0.61	NR	0.47	NR	0.86

FA = formaldehyde; SD = standard deviation; NR = not reported

[a] Numbers given after a forward slash indicate concentrations with four 15-minute peaks.

Benchmark Dose-response Modeling

Table 26 presents the BMD modelling results of the mean change in eye blinking frequency at end of exposure compared to pre-exposure at concentrations without peaks, by sensitivity group. Models in hyposensitives and hypersensitives yielded BMDs and BMDLs higher than the maximum concentration suggesting caution in using such result for POD derivation.

Table 26. Benchmark dose modeling for tear film break-up time.

Sensitivity Group	Model	BMD (ppm)	BMDL (ppm)	Notes
Hypo	Linear (non-constant variance)	2.225262	0.965142	Lowest AIC; BMD higher than maximum dose, BMDL higher than maximum dose
Hyper	Polynomial Degree 2	2.151421	0.92565	Lowest AIC; BMD higher than maximum dose, BMDL higher than maximum dose

BMD = benchmark dose; BMDL = benchmark dose level; AIC = Akaike information criterion

Nasal Flow

Mueller at el. (2013) observed a significantly increased mean nasal flow in hypersensitives at 0.7 ppm compared to pre-exposure values. Significant differences were observed in the responses of hyposensitives at 0.3 ppm with peaks at 0.6 ppm compared to 0 ppm controls. The patterns were not consistent. Results of exposure to formaldehyde on changes in nasal flow are shown in **Table 27** (using data from Table Online Resource 8 of Mueller et al. (2008)). Our analyses confirmed the previously reported significant findings.

Table 27. Summary of results for nasal flow.

FA, ppm	Sensitivity Group	Mean Difference (SD)	p-values compared to pre-exposure		p-values compared to control condition	
			Published	*ICF*	*Published*	*ICF*
0	Hypo	−50.30 (314.09)	NR	0.48	–	–
	Hyper	38.48 (213.51)	NR	0.41	–	–
0.3/0.6[a]	Hypo	157.10 (339.48)	NR	<0.05	<0.05	0.05

FA, ppm	Sensitivity Group	Mean Difference (SD)	p-values compared to pre-exposure		p-values compared to control condition	
			Published	*ICF*	*Published*	*ICF*
0.4/0.8	Hyper	−10.67 (334.40)	NR	0.88	NR	0.57
	Hypo	−7.05 (303.20)	NR	0.92	NR	0.66
	Hyper	50.57 (346.11)	NR	0.51	NR	0.89
0.5	Hypo	−18.55 (406.98)	NR	0.84	NR	0.78
	Hyper	−77.81 (306.18)	NR	0.25	NR	0.16
0.7	Hypo	83.00 (266.55)	NR	0.17	NR	0.16
	Hyper	192.76 (305.88)	<0.01	<0.01	NR	0.07

FA = formaldehyde; SD = standard deviation; NR = not reported

[a] Numbers given after a forward slash indicate concentrations with four 15-minute peaks.

Benchmark Dose-response Modeling

Table 28 presents the BMD modelling results of the mean change in nasal flow at end of exposure compared to pre-exposure at concentrations without peaks, by sensitivity group. Models in hyposensitives and hypersensitives yielded BMDs of 1.66 ppm and 0.71 ppm, respectively, higher than the maximum concentration suggesting caution in using such result for POD derivation. BMDLs were 0.98 and 0.68 ppm, in hyposensitives and hypersensitives respectively.

Table 28. Benchmark dose modeling for nasal flow.

Sensitivity Group	Model	BMD (ppm)	BMDL (ppm)	Notes
Hypo	Polynomial Degree 2	1.661333	0.983554	Lowest AIC; BMD higher than maximum dose, BMDL higher than maximum dose
Hyper	Power	0.710256	0.682046	Lowest AIC; BMD higher than maximum dose

BMD = benchmark dose; BMDL = benchmark dose level; AIC = Akaike information criterion.

Subjective Symptoms

Total Symptom Score

Mueller at el. (2013) observed a significantly increased changes in cumulative subjective symptoms in hypersensitives at 0.3 ppm with 0.6 ppm peaks and at 0.4 ppm with 0.8 ppm peaks, compared to pre-exposure values. Results of exposure to formaldehyde on changes in total symptom score are shown in **Table 29** (using data from Table Online Resource 9 of Mueller et al. (2008)). Our analyses confirmed the previously reported significant findings.

Table 29. Summary of results for total symptom score.

FA, ppm	Sensitivity Group	Mean Difference (SD)	p-values compared to pre-exposure	
			Published	*ICF*
0	Hypo	−0.20 (2.44)	NR	0.72
	Hyper	0.97 (3.85)	NR	0.26
0.3/0.6[a]	Hypo	0.57 (1.37)	NR	0.07
	Hyper	2.20 (2.24)	<0.001	<0.001
0.4/0.8	Hypo	0.59 (2.63)	NR	0.32
	Hyper	2.03 (2.80)	<0.01	<0.01
0.5	Hypo	0.63 (1.88)	NR	0.14
	Hyper	0.84 (3.87)	NR	0.33
0.7	Hypo	0.68 (1.59)	NR	0.06
	Hyper	1.63 (4.77)	NR	0.13

FA = formaldehyde; SD = standard deviation; NR = not reported

[a] Numbers given after a forward slash indicate concentrations with four 15-minute peaks.

Benchmark Dose-response Modeling

Table 30 presents the BMD modelling results of the total symptom score at end of exposure compared to pre-exposure at concentrations without peaks, by sensitivity group. Models in hyposensitives yielded higher BMD and BMDL than in hypersensitives, and higher than the maximum concentration suggesting caution in using such result for POD derivation. BMDLs were 1.02 and 0.71 ppm, in hyposensitives and hypersensitives respectively.

Table 30. Benchmark dose modeling for total symptom score.

Sensitivity Group	Model	BMD (ppm)	BMDL (ppm)	Notes
Hypo	Linear	1.82919	1.018144	Lowest AIC; BMD higher than maximum dose, BMDL higher than maximum dose
Hyper	Exponential 3	0.742646	0.706038	Lowest AIC; BMD higher than maximum dose, BMDL higher than maximum dose

BMD = benchmark dose; BMDL = benchmark dose level; AIC = Akaike information criterion

Eye Irritation

Mueller at el. (2013) observed increased eye irritation subscores after exposure in most groups, but no significant changes. Results of exposure to formaldehyde on changes in eye irritation are shown in **Table 31** (using data from Table Online Resource 10 in Mueller et al. (2008)). Our analyses confirmed the reported lack of significant findings, except for a significant increase in hypersensitives exposed to 0.3 ppm with 0.6 ppm peaks, and a borderline significant increase (p-value =0.05) in hypersensitives exposed to 0.5 ppm.

Table 31. Summary of results for eye irritation.

			p-values compared to pre-exposure	
FA, ppm	**Sensitivity Group**	**Mean Difference (SD)**	*Published*	*ICF*
0	Hypo	−0.17 (2.02)	NR	0.71
	Hyper	1.96 (7.59)	NR	0.24
0.3/0.6[a]	Hypo	0.23 (2.65)	NR	0.70
	Hyper	2.13 (4.71)	NR	<0.05
0.4/0.8	Hypo	0.62 (5.71)	NR	0.63
	Hyper	1.43 (5.31)	NR	0.22
0.5	Hypo	−0.09 (2.14)	NR	0.85
	Hyper	1.24 (2.84)	NR	0.05
0.7	Hypo	0.94 (4.56)	NR	0.36
	Hyper	0.52 (4.14)	NR	0.57

FA = formaldehyde; SD = standard deviation; NR = not reported

[a] Numbers given after a forward slash indicate concentrations with four 15-minute peaks.

Benchmark Dose-response Modeling

Table 32 presents the BMD modelling results of the mean change in eye irritation at end of exposure compared to pre-exposure at concentrations without peaks, by sensitivity group. No model was viable for the data in hyposensitive volunteers. The estimated BMD and BMDLs in hypersensitive volunteers are 8.38 ppm and 0.75 ppm, respectively, higher than the maximum concentration suggesting caution in using such result for POD derivation.

Table 32. Benchmark dose modeling for eye irritation.

Sensitivity Group	Model	BMD (ppm)	BMDL (ppm)	Notes
Hypo	No viable models	–	–	
Hyper	Power (non-constant variance)	0.829367	0.754982	Lowest AIC; BMD higher than maximum dose, BMDL higher than maximum dose

BMD = benchmark dose; BMDL = benchmark dose level; AIC = Akaike information criterion

Nasal Irritation

Mueller at el. (2013) observed increased and decreased nasal irritation after exposure in most groups, but no significant changes. Results of exposure to formaldehyde on changes in nasal irritation are shown in **Table 33** (using data from Table Online Resource 11 in Mueller et al. (2008)). Our analyses confirmed the reported lack of significant findings.

Table 33. Summary of results for nasal irritation.

			p-values compared to pre-exposure	
FA, ppm	**Sensitivity Group**	**Mean Difference (SD)**	*Published*	*ICF*
0	Hypo	−1.13 (4.12)	NR	0.23
	Hyper	−0.09 (1.60)	NR	0.80
0.3/0.6[a]	Hypo	0.20 (1.84)	NR	0.63
	Hyper	0.67 (2.62)	NR	0.25
0.4/0.8	Hypo	−0.56 (4.85)	NR	0.61
	Hyper	−0.41 (3.64)	NR	0.61
0.5	Hypo	0.73 (5.29)	NR	0.54
	Hyper	−0.74 (4.86)	NR	0.49
0.7	Hypo	−0.97 (4.56)	NR	0.35
	Hyper	−0.71 (8.56)	NR	0.71

FA = formaldehyde; SD = standard deviation; NR = not reported

[a] Numbers given after a forward slash indicate concentrations with four 15-minute peaks.

Benchmark Dose-response Modeling

Table 34 presents the BMD modelling results of the nasal irritation score at end of exposure compared to pre-exposure at concentrations without peaks, by sensitivity group. Models in hyposensitives yielded higher BMD and BMDL than in hypersensitives, and higher than the maximum concentration suggesting caution in using such result for POD derivation. BMD (BMDLs) were 2.56 (1.10) ppm and 0.94 (0.77) ppm, in hyposensitives and hypersensitives respectively.

Table 34. Benchmark dose modeling for nasal irritation.

Sensitivity Group	Model	BMD (ppm)	BMDL (ppm)	Notes
Hypo	Polynomial Degree 2	2.563198	1.103389	Lowest AIC; BMD higher than maximum dose, BMDL higher than maximum dose
Hyper	Polynomial Degree 2 (non-constant variance)	0.944075	0.773216	Lowest AIC; BMD higher than maximum dose, BMDL higher than maximum dose

BMD = benchmark dose; BMDL = benchmark dose level; AIC = Akaike information criterion

Olfactory Symptoms

Mueller et al. (2013) reported increased olfactory symptoms compared to the control at each concentration. The increase was statistically significant for hypersensitive volunteers exposed at 0.4 ppm with peaks at 0.8 ppm. Hypersensitive volunteers reported consistently higher complaints than hyposensitive volunteers, statistically significantly higher at 0.3 ppm with 0.6 ppm peaks and at 0.4 ppm with 0.8 ppm peaks. Results of exposure to formaldehyde on changes in olfactory symptoms are shown

in in **Table 35** (using data from Table Online Resource 12 in Mueller et al. (2013)). Mean differences between post- and pre-exposure symptoms were consistently positive, with statistically significant differences observed at all concentrations above the control condition for the hypersensitive group, and at 0.4 ppm with 0.8 ppm peaks and 0.5 ppm in the hyposensitive group. Our analyses confirmed the significant findings when comparing hypersensitive volunteers with hyposensitive volunteers at concentrations with peaks. When comparing to control condition, no significant changes were observed in our analyses.

Table 35. Summary of results for olfactory symptoms.

FA, ppm	Sensitivity Group	Mean Difference (SD)	p-values compared to pre-exposure		p-values compared to control condition		p-values compared to hyposensitives	
			Published	*ICF*	*Published*	*ICF*	*Published*	*ICF*
0	Hypo	0.10 (4.05)	NR	0.91	–	–	–	–
	Hyper	4.98 (11.11)	NR	<0.05	–	–	NR	0.07
0.3/0.6[a]	Hypo	2.49 (5.84)	NR	0.06	NR	0.14	–	–
	Hyper	8.43 (8.28)	<0.001	<0.001	NR	0.26	<0.05	0.01
0.4/0.8	Hypo	3.73 (7.29)	<0.05	<0.05	NR	0.06	–	–
	Hyper	11.60 (14.25)	<0.01	<0.001	<0.05	0.10	<0.01	0.03
0.5	Hypo	2.04 (2.36)	<0.01	<0.001	NR	0.07	–	–
	Hyper	6.23 (9.01)	<0.01	<0.01	NR	0.69	NR	0.05
0.7	Hypo	2.71 (7.14)	NR	0.10	NR	0.16	–	–
	Hyper	7.31 (15.57)	<0.05	<0.05	NR	0.58	NR	0.23

FA = formaldehyde; SD = standard deviation; NR = not reported

[a] Numbers given after a forward slash indicate concentrations with four 15-minute peaks.

Benchmark Dose-response Modeling

Table 36 presents the BMD modelling results of the mean change in olfactory symptoms at end of exposure compared to pre-exposure at concentrations without peaks, by sensitivity group. No model was viable for the data in hyposensitive volunteers. The estimated BMD and BMDLs in hypersensitive volunteers are 0.76 ppm and 0.71 ppm, respectively, higher than the maximum concentration suggesting caution in using such result for POD derivation.

Table 36. Benchmark dose modeling for olfactory symptoms.

Sensitivity Group	Model	BMD (ppm)	BMDL (ppm)	Notes
Hypo	No viable models	–	–	
Hyper	Power (non-constant variance)	0.759959	0.707109	Lowest AIC; BMD higher than maximum dose, BMDL higher than maximum dose

BMD = benchmark dose; BMDL = benchmark dose level; AIC = Akaike information criterion

Perception of Impure Air

Mueller et al. (2013) observed consistently higher postexposure values, with these mean differences reaching statistical significance for hypersensitives at all concentrations (including the control condition) and for hyposensitives at 0.4 ppm and 0.5 ppm, but not when comparing to control. Results of exposure to formaldehyde on changes in perception of impure air are shown in in **Table 37**, (using data from Table Online Resource 13 in Mueller et al. (2013)). Our analyses confirmed the reported significant changes between post- and pre-exposure, and between hypo- and hypersensitive volunteers.

Table 37. Summary of results for perception of impure air.

FA, ppm	Sensitivity Group	Mean Difference (SD)	p-values compared to pre-exposure		p-values compared to hyposensitives	
			Published	*ICF*	*Published*	*ICF*
0	Hypo	1.15 (8.53)	NR	0.55	–	–
	Hyper	17.62 (22.98)	<0.01	<0.01	<0.01	<0.01
0.3/0.6[a]	Hypo	4.35 (10.07)	NR	0.06	–	–
	Hyper	23.90 (24.55)	<0.001	<0.001	<0.01	<0.01
0.4/0.8	Hypo	7.35 (13.35)	<0.05	0.02	–	–
	Hyper	29.52 (26.37)	<0.001	<0.001	<0.001	<0.01
0.5	Hypo	4.60 (5.83)	<0.01	<0.01	–	–
	Hyper	21.38 (25.02)	<0.001	<0.001	<0.01	<0.01
0.7	Hypo	6.75 (22.17)	NR	0.18	–	–
	Hyper	24.95 (25.41)	<0.001	<0.001	<0.01	0.02

FA = formaldehyde; SD = standard deviation; NR = not reported

[a] Numbers given after a forward slash indicate concentrations with four 15-minute peaks.

Benchmark Dose-response Modeling

Table 38 presents the BMD modelling results of the perception of impure air at end of exposure compared to pre-exposure at concentrations without peaks, by sensitivity group. No model was viable for the data in hyposensitive volunteers. The estimated BMD and BMDLs in hypersensitive volunteers are 2.42 ppm and 0.89 ppm, respectively, higher than the maximum concentration suggesting caution in using such result for POD derivation.

Table 38. Benchmark dose modeling for perception of impure air.

Sensitivity Group	Model	BMD (ppm)	BMDL (ppm)	Notes
Hypo	No viable models	–	–	–
Hyper	Linear	2.4166	0.89221	Lowest AIC; BMD and BMDL higher than maximum dose

BMD = benchmark dose; BMDL = benchmark dose level; AIC = Akaike information criterion

References

EPA (2012). Benchmark Dose Technical Guidance. Risk assessment Forum, US Environmental Protection Agency. EPA/100/R-12/001. June 2012.

EPA (2022a). Toxicological Review of Formaldehyde – Inhalation. [CASRN 50-00-0]. External Review Draft. EPA/635/R-22/039a. April 2022.

EPA (2022b). Benchmark Dose Software (BMDS) (Build 3.3; Model Library Version 2022.10) [Computer Software]. Available from https://www.epa.gov/bmds/download-bmds.

Lang I, Bruckner T, Triebig G. Formaldehyde and chemosensory irritation in humans: a controlled human exposure study. Regul Toxicol Pharmacol. 2008 Feb;50(1):23-36. doi: 10.1016/j.yrtph.2007.08.012. PMID: 17942205.

Mueller JU, Bruckner T, Triebig G. Exposure study to examine chemosensory effects of formaldehyde on hyposensitive and hypersensitive males. Int Arch Occup Environ Health. 2013 Jan;86(1):107-17. doi: 10.1007/s00420-012-0745-9. PMID: 22371090.

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL

Rosner, Bernard A. (2015). Fundamentals of biostatistics. 8th Edition. Boston: Brooks/Cole, Cengage Learning.

SAS Institute Inc 2013. SAS/ACCESS® 9.4. Cary, NC: SAS Institute Inc.

Appendices

Appendix 1. Notes on formaldehyde papers by Lang et al. (2008) and Mueller et al. (2013)

Appendix 2. Excel file with data used for reanalysis and summary results for Lang et al (2008) and Mueller et al. (2008) (Lang_Mueller_FA_DataResults.xlxs)

Appendix 3. Code and output from statistical analyses in SAS and R.

Appendix 4. BMDS modelling Excel outputs.

Appendix 1. Notes on formaldehyde papers by Lang et al. 2008 and Mueller et al. 2013

Jonathan Cohen, Nov 4, 2022.

Notes on formaldehyde papers by Lang et al. 2008 and Mueller et al. 2013

Lang et al. 2008

21 subjects were each measured for 4 hours under 10 formaldehyde exposure (ppm) conditions in a randomized order: 0 ppm, 0.15 ppm, 0.3 ppm, 0.3 ppm with 4 peaks at 0.6 ppm, 0.5 ppm, 0.5 ppm with 4 peaks at 1.0 ppm, 0 ppm with 12-16 ppm of a masking agent ethyl acetate (EA), 0.3 ppm with 12-16 ppm EA, 0.5 ppm with 12-16 ppm EEA, 0.5 ppm with 4 peaks at 1.0 ppm and 12-16 ppm EA.

Included in the IRIS report but not selected because "difficult to define adverse level cut-off for these endpoints [eye irritation]." This issue is a concern if ICF tries to estimate BMDLs.

As described in detail below, without obtaining the raw data or approximating data values from the graphs, ICF's proposed statistical analyses would require assuming independence between responses at different doses and will be limited to analyses of conjunctival redness, blinking frequency, and the Spearman correlation between blinking frequency and eye irritation at 0.5 ppm with peaks.

Gender

Authors: Found no gender-related statistically significant differences except for eye redness at 2 formaldehyde levels and ocular irritation. Data not shown.

ICF: Without data, no statistical analyses are feasible.

Conjunctival redness

Authors: McNemar test to compare rates of moderate redness at 0 ppm with other levels at 195 mins (statistically significant increase at 0.5 ppm with 4 peaks at 1.0 ppm without EA) and to compare results at other times of day (no statistically significant influence). Graphed rates at different time points. Cannot replicate these analyses without the raw data.

ICF: Without raw data we must assume independence between responses at different doses. In that case we can compare rates for pairs of doses using Fisher's Exact test, evaluate trends using the Cochran-Armitage trend test, fit dose-response models, and compute the BMD for a selected BMR, e.g., a rate of 10% above the control rate. For the trend analyses and dose-response modeling can analyze trends with and without EA. To include the results with formaldehyde peaks we can calculate the time-averaged formaldehyde assuming that each peak was for 15 minutes. The lengths of the peaks were not stated in the Lang et al. paper, but their colleagues wrote in the Mueller et al. paper that the Lang et al. peaks were for 15 minutes.

Blinking Frequency (BF)

Authors: They found no increases in BF with time; data not provided. Compared mean BF for each dose versus control using a repeated measures two-way ANOVA with contrasts and found statistically significant differences at the dose 0.5 ppm with 4 peaks at 1.0 ppm, with or without EA. They also mention using Kruskal-Wallis non-parametric tests in cases of non-homogeneity, although these are not

mentioned in their Table 7. They tabulated the means, standard deviations, medians, and ranges. Cannot replicate these analyses without the raw data.

ICF: Without the raw data we must assume independence between responses at different doses. In that case we can compare pairs of doses using a one-way ANOVA with contrasts, fit dose-response models, and compute the BMD for a selected BMR, e.g., a mean 50% above the control mean. For the trend analyses and dose-response modeling we can analyze trends with and without EA.

Nasal Resistance and Flow

Authors: Differences after exposure compared to before exposure were not statistically significant and occurred for the control doses, and thus were assumed to be unrelated to the formaldehyde or EA exposure. No data provided.

ICF: Without data, no statistical analyses are feasible.

Pulmonary function

*Au*thors: Differences between day 1 and day 10 were not statistically significant. No data provided.

ICF: Without data, no statistical analyses are feasible.

Reaction times

Authors: Visual stimulus reaction time increased significantly at 0.3 ppm with or without EA. Acoustic stimulus reaction time increased significantly at 0.3 ppm without EA. Combined visual and acoustic reaction time increased significantly at 0.3 ppm without EA. Results displayed as box-whisker plots. Statistical analyses used a repeated measures two-way ANOVA with contrasts. Also found no changes in motor reaction times, but data not shown. As those effects were not observed at higher doses and the changes were "slight", these findings were considered to be incidental.

ICF: Without raw data we must assume independence between responses at different doses. In addition, since the data are only reported in box-whisker plots, a statistical analysis would require estimating the means and standard deviations from the box-whisker plots, assuming normal distributions. Since this approach would be quite inaccurate, it is not recommended. However, if we do want to approximate the statistics from the box-whisker plots, then we can compare pairs of doses using a one-way ANOVA with contrasts, fit dose-response models, and compute the BMD for a selected BMR, e.g., a mean 50% above the control mean. For the trend analyses and dose-response modeling can analyze trends with and without EA. To include the results with formaldehyde peaks we can calculate the time-averaged formaldehyde assuming that each peak was for 15 minutes. The lengths of the peaks were not stated in the Lang et al. paper, but their colleagues wrote in the Mueller et al. paper that the Lang et al. peaks were for 15 minutes.

Subjective ratings

Authors: Subjective ratings for the total symptom score are reported in box-and whisker plots, and the means and standard deviations of symptom scores are reported in line plots for eye irritation, nasal irritation, olfactory symptoms, and respiratory irritation. The mean total symptom score was significantly higher versus the controls at 0.5 ppm with peaks, with or without EA. In most cases the

mean scores for eye irritation, nasal irritation, olfactory symptoms, respiratory irritation, and annoyance were significantly higher than the controls.

ICF: Without raw data we must assume independence between responses at different doses. Same suggested approach and caveats as for Reaction Times above.

Personality factors

Authors: Added negative affinity as a covariate to the statistical models and found that several statistical findings for the subjective scores became no longer significant.

ICF: Without data, no statistical analyses are feasible.

Correlations between subjective eye irritation and blinking frequency

Authors: Created scatter plots of blinking frequency versus eye irritation for formaldehyde at 0.5 ppm with peaks, with or without EA. After removing an outlier, found a significant correlation with EA but not without EA. It is not clear from the text whether they used Pearson or Spearman rank correlations.

ICF: Should be possible to calculate the Spearman rank correlations and their statistical significance using the scatter plots.

Mueller et al. 2013

41 male subjects were each measured for 4 hours under 5 formaldehyde exposure (ppm) conditions in a randomized order: 0 ppm, 0.3 ppm with 4 peaks for 15 minutes each at 0.6 ppm, 0.4 ppm with 4 peaks for 15 minutes each at 0.8 ppm, 0.5 ppm, 0.7 ppm. The subjects were divided into two groups of hypo- or hyper-sensitive individuals based on the mean measured 'sensation of pain' caused by a nasally applied CO_2 gas and air mixture on a visual analogue scale (VAS). The VAS scale ranged from "none at all" (expressed as 0 mm) to "unbearable" (expressed as 100 mm). The mean scores from 3 measurements each at doses of 40, 60, or 80 percent were added to give the individual CO_2 sensitivity score (0-300 mm), and the 20 hypo- and 21 hyper-sensitive individuals were the subjects with CO_2 sensitivity scores below or above the median of 80.3 mm. No measurements were made during the exposure except for Blinking Frequency and the "SPES" questionnaire on symptoms and complaints that were both measured in the last 15 minutes of exposure.

Included in IRIS report but not selected because "An exposure-response trend was not observed for either endpoint. Difficult to define an adverse response level cutoff for these endpoints." The endpoints considered in IRIS were for eye irritation, measured by the tear film break-up time and by the symptom score. This issue is a concern if ICF tries to estimate BMDLs.

As discussed below, in the absence of the raw data, ICF's proposed statistical analysis approach would require assuming independence between responses at different doses. Under that assumption we can use the tabulated data in the Online Resources to reanalyze each of the responses other than conjunctival redness using a) one-sample t tests for mean differences after exposure compared to pre-exposure, b) one-way ANOVA to compare pairs of doses, c) fit dose-response models to estimate trends, d) estimate BMDs for a selected BMR, e) test for differences between hypo- and hyper-sensitive males by adding a sensitivity group covariate to the ANOVA models. Analyses of conjunctival redness is not feasible without an explanation of the percentages in Online Resource 3.

To use the Online Resources to re-analyze the data it will be necessary to interpret how the outliers were treated, since they frequently refer to "omitted outlier values" and state that "scaling was adapted and values outside the selected scale were omitted (marked by open squares)". The text makes it clear that extremely high or low values in the *plots* were truncated, but it is not clear from the paper whether the *tabulated* means and standard deviations in the Online Resources use all the data or use the data after either excluding the outliers or truncating them to the given scales. It is hard to tell what was done from the graphs and tables. If we can contact the authors for more details, then we can address these issues directly, but otherwise we will assume that the tabulated data were calculated from the complete data without excluding outliers, and the sample sizes were 20 hypo- and 21 hyper-sensitive individuals. Some of the findings might be impacted by the outliers.

Conjunctival redness

Authors: See Online Resources 3 to 5. Reported as statistically significantly increased redness after exposure for hyposensitives at 0 ppm, and statistically significantly decreased redness after exposure for hypersensitives at 0 ppm. It is not at all clear from the text and the tabulated values how this analysis was done, since the tabulated numbers seem to be the percentages of subjects where the redness either decreased, remained constant, or increased, rather than the percentage changes in the average redness. It is also unclear why the same percentage decrease of 23.81% for hypersensitives is significant at a dose of 0 ppm but not at 0.4 ppm with 4 peaks at 0.8 ppm.

ICF: In the absence of an explanation for the data in Table Online Resource 3, it is not feasible to check this analysis.

Blinking Frequency (BF)

Authors: They plotted mean differences after exposure for hypo- and hyper-sensitives at each dose in box-whisker plots. Based on a one-sample t test, there was a statistically significant change for hypersensitives at a dose of 0 ppm. In general, they found decreases after exposure, but they found no consistent statistically significant change. Differences between hypo- and hyper-sensitives were not statistically significant based on ANCOVA. A consistent dose-effect relationship was not found.

ICF: Means and standard deviations of the differences after exposure by dose and sensitivity group are tabulated in Online Resource 6. In the absence of the raw data, we will need to assume independence between responses at different doses. We can apply one-sample t tests for the mean differences for each dose and sensitivity group. For each sensitivity group, we can compare pairs of doses using a one-way ANOVA with contrasts, fit dose-response models to estimate trends, and compute the BMD for a selected BMR, e.g., a specified mean difference. For the trend analyses and dose-response modeling, we can include the results with formaldehyde peaks based on the time-averaged formaldehyde exposure. We can also test for differences between sensitivity groups by fitting a model with terms for each dose and a sensitivity group covariate.

Self-reported Tear Film Breakup Time (sBUT)

Authors: They plotted mean differences after exposure for hypo- and hyper-sensitives at each dose in box-whisker plots. Based on one-sample t tests, there were several statistically significant changes for different doses and sensitivity groups. There were also several statistically significant differences between doses and between hypo- and hyper-sensitives. A consistent dose-effect relationship was not found.

ICF: Means and standard deviations of the differences after exposure by dose and sensitivity group are tabulated in Online Resource 7. Various possible statistical analyses are described above under "Blinking Frequency."

Nasal Flow

Authors: They plotted mean differences after exposure for hypo- and hyper-sensitives at each dose in box-whisker plots. Based on one-sample t tests, there was a statistically significant change in the mean for hypersensitives at 0.7 ppm. Based on ANCOVA, there was a statistically significant difference in the responses of hyposensitives at 0 ppm versus 0.3 ppm with peaks at 0.6 ppm. The patterns were not consistent.

ICF: Means and standard deviations of the differences after exposure by dose and sensitivity group are tabulated in Online Resource 8. Various possible statistical analyses are described above under "Blinking Frequency."

SPES Sum Score

Authors: They plotted mean differences after exposure for hypo- and hyper-sensitives at each dose in box-whisker plots. The mean scores increased after exposure in all but one case (hyposensitives at dose 0). Based on one-sample t tests, there was a statistically significant change in the mean for hypersensitives at 0.3 ppm with peaks at 0.6 ppm and at 0.4 ppm with peaks at 0.8 ppm. Mean differences were consistently higher for hypersensitives versus hyposensitives.

ICF: Means and standard deviations of the differences after exposure by dose and sensitivity group are tabulated in Online Resource 9. Various possible statistical analyses are described above under "Blinking Frequency."

SPES Eye Irritation Subscore

Authors: In Online Resource 10, they plotted mean differences after exposure for hypo- and hyper-sensitives at each dose in box-whisker plots. The mean scores generally increased after exposure. Based on one-sample t tests, there were no statistically significant changes in the mean. A concentration-effect relationship was not found.

ICF: Means and standard deviations of the differences after exposure by dose and sensitivity group are tabulated in Online Resource 10. Various possible statistical analyses are described above under "Blinking Frequency."

SPES Nasal Irritation Subscore

Authors: In Online Resource 11, they plotted mean differences after exposure for hypo- and hyper-sensitives at each dose in box-whisker plots. The mean scores both increased and decreased after

exposure for different dose and sensitivity group combinations. Based on one-sample t tests, there were no statistically significant changes in the mean. A concentration-effect relationship was not found.

ICF: Means and standard deviations of the differences after exposure by dose and sensitivity group are tabulated in Online Resource 11. Various possible statistical analyses are described above under "Blinking Frequency."

SPES Olfactory Symptom Subscore

Authors: They plotted mean differences after exposure for hypo- and hyper-sensitives at each dose in box-whisker plots. The mean scores generally increased after exposure. Based on one-sample t tests, there were several statistically significant changes in the mean. Increased symptoms compared to the control dose were found at each of the positive concentrations, and the increase was statistically significant for hypersensitive volunteers exposed at 0.4 ppm with peaks at 0.8 ppm. A concentration-effect relationship was not found. Hypersensitive volunteers reported consistently higher complaints than hyposensitive volunteers, statistically significantly higher at 0.3 ppm with 0.6 ppm peaks and at 0.4 ppm with 0.8 ppm peaks.

ICF: Means and standard deviations of the differences after exposure by dose and sensitivity group are tabulated in Online Resource 12. Various possible statistical analyses are described above under "Blinking Frequency."

SPES Perception of Pure Air Subscore

Authors: They plotted mean differences after exposure for hypo- and hyper-sensitives at each dose in box-whisker plots. The mean scores consistently increased after exposure. Based on one-sample t tests, there were statistically significant changes in the mean at every dose (including 0 ppm) for hypersensitives, and at doses of 0.3 ppm with 0.6 ppm peaks and 0.5 ppm for hyposensitives. A concentration-effect relationship was not found.

ICF: Means and standard deviations of the differences after exposure by dose and sensitivity group are tabulated in Online Resource 13. Various possible statistical analyses are described above under "Blinking Frequency."

Positive and Negative Affect Schedule (PANAS)

Authors: The PANAS scores measure personality patterns. Supplementary analyses of the SPES scores with a PANAS negative affectivity covariate found no significant influence on the SPES scores. The detailed results were not reported.

ICF: In the absence of the raw data, it is not possible to validate those findings.

Appendix 2. Excel file with data used for reanalysis and summary results for Lang et al (2008) and Mueller et al. (2008)

Separate file: Lang_Mueller_FA_DataResults.xlxs

Appendix 3. Code and output from statistical analyses in SAS and R.

Separate zipped file including:

Lang 2008 R Code.txt,
Mueller 2013 R Code.txt,
Lang_FisherCochran_Corr.sas
Lang_Fishercochran_Corr_Output.pdf

Appendix 4. BMDS modelling Excel outputs.

Separate zipped file with BMDS detailed results: Appendix4_Lang_Mueller_FA_BMD