Office of Environmental Health Hazard Assessment

Joan E. Denton, Ph.D., Director
Headquarters • 1001 I Street • Sacramento, California 95814
Mailing Address: P.O. Box 4010 • Sacramento, California 95812-4010
Oakland Office • Mailing Address: 1515 Clay Street, 16th Floor • Oakland, California 94612

Linda S. Adams
Secretary for Environmental Protection

Arnold Schwarzenegger
Governor

MEMORANDUM

TO: Dr. Dennis Shusterman
Hazard Evaluation System and Information Service (HESIS)
850 Marina Bay Parkway, Bldg P, 3rd Floor
Richmond, California 94804

FROM: Sara Hoover, M.S.
Chief, Safer Alternatives Assessment and Biomonitoring Section
Office of Environmental Health Hazard Assessment

DATE: June 19, 2009

SUBJECT: UPDATED BENCHMARK CONCENTRATION ANALYSIS FOR N-METHYLPYRROLIDONE AND DISCUSSION OF COMMENTS

In response to a request by the Hazard Evaluation System and Information Service (HESIS), the Office of Environmental Health Hazard Assessment (OEHHA) conducted a benchmark concentration (BMC) analysis on N-methylpyrrolidone (NMP) (OEHHA, 2009). We provided the results of the BMC analysis to HESIS in a memorandum dated March 5, 2009. Comments on the analysis were submitted to the Department of Occupational Safety and Health (DOSH) by the NMP Producers Group and The Sapphire Group. Below is a discussion of key points raised in the comments and a description of updates OEHHA has made to the BMC analysis based on the comments.

The NMP Producers Group also submitted to DOSH an unpublished manuscript that describes a physiologically-based pharmacokinetic model developed for NMP (Poet *et al.*, 2009). OEHHA requested and obtained additional details on the model and is currently reviewing it.

Concurrent control group for Staples (1990)

The Sapphire Group identified a second control group in the Staples (1990) study that they recommend including as concurrent controls. OEHHA re-examined the Staples study and agrees with The Sapphire Group that part of one control group (CI) and all of a second control group (CII) can be considered concurrent controls for the analysis of the pup weight data. The CI control group included a total of 40 females at the beginning of the study while the CII control group included 20 females. Twenty females from CI and 20 females from CII were designated as being allowed to give birth, which would yield pup weight measurements. An additional 20 females in control group CI were killed at gestation day (gd) 21 for examination of fetal weight and developmental abnormalities. OEHHA agrees it is appropriate to use half of CI and all of CII as concurrent controls in the analysis of pup weight data from Staples (1990). The Staples concurrent control group data are summarized below.

	Pup bodyweight post-partum day one		
Staples control group	***Mean (g)***	***St dev (g)***	***N***
CI (half)	7.47	0.76	20
CII	7.49	0.65	19
Combined	7.48	0.70	39

Nominal versus measured concentrations

OEHHA agrees with The Sapphire Group that the measured air concentration is the most appropriate metric for the benchmark concentration analysis. Both Staples (1990) and Saillenfait *et al.* (2003) tabulated individual dam data using the nominal (*i.e.*, target) concentrations. In most cases, the nominal and measured concentrations are very similar and often within rounding. However, for the Staples (1990) study, there was a clear difference between the nominal and measured concentrations at the highest exposure. The nominal concentrations for the Staples study were 0, 10, 50, and 130 ppm, while the measured concentrations were 0, 10.3, 50.8 and 116 ppm. For the Saillenfait *et al.* (2003) study, the nominal concentrations were 0, 30, 60, and 120 ppm, while the measured concentrations were 0, 30, 60, and 121 ppm. The use of measured concentrations produces a BMCL of 43 ppm for Staples, compared to a BMCL of 48 ppm using the nominal concentrations (both of these analyses used the larger concurrent control group as discussed above for Staples).

Combining multiple exposure groups in Staples (1990)

The Sapphire Group suggested pooling the pup weight results from two high exposure groups (EIII and EIV) reported by Staples (1990). However, these two exposure groups are not comparable and Staples (1990) reported them as separate experiments. The Staples (1990) experimental design involved exposing both males (bucks) and females (dams) to 0, 10, 50, and 130 ppm NMP for 12 weeks prior to mating and during pregnancy (dams only). The 130 ppm exposure group in this part of the experiment was designated by Staples as EIII. In separate experiments, Staples exposed only the dams to 130 ppm prior to mating and during pregnancy (Group EIV) or only the bucks to 130 ppm prior to mating (Group EV). When conducting trend analyses Staples only included the exposure groups within the same experimental design, explicitly excluding Groups EIV and EV.

The data in the table below show pup body weights for the control group and all three high exposure groups (EIII, EIV, and EV) nominally exposed to 130 ppm (actual concentration for all three high exposure groups was measured to be 116 ppm). Pups from the EIII group, in which both parents were exposed, had statistically significantly lower body weights compared to controls (Staples, 1990). Though not statistically significant, there was also a reduction in the body weight of pups from group EIV (only dams exposed) and EV (only bucks exposed) at day one post-partum compared to concurrent controls. These pup body weight differences were noted by Staples (1990): "The pups born to the groups in which only one of the parents received NMP at the [130ppm] exposure level also weighed less than the control pups." OEHHA recognizes that because these changes are not statistically significant, the differences could be due to chance. However, these data are consistent with the interpretation that exposure of the males prior to mating and/or exposure of the females prior to mating and during pregnancy could impact the body weight of the offspring. Not only is it generally recognized that developmental toxicity can result from "exposure prior to conception (either parent)" (U.S. EPA, 1991), there is mounting evidence that paternal as well as maternal exposures can potentially influence offspring development through, for example, epigenetic mechanisms. Further, it is apparent that the combined effect of both parents being exposed to NMP results in a more pronounced body weight reduction compared to exposure of only one parent. These findings highlight two things: (1) exposure 12-weeks prior to mating (included in the Staples design and not in the Saillenfait design) has a potential influence on the pup body weight in addition to the effect of exposing the fetus during pregnancy; and (2) the EIII, EIV and EV exposure groups in Staples (1990) are

distinct so exposure groups EIII and EIV should not be combined for benchmark concentration analysis.

Staples (1990) exposure group	***Pup bodyweight post-partum day one***		
	Mean (g)	***St dev (g)***	***N***
Control	7.48	0.70	39
EV (males only)	7.29	0.65	17
EIV (females only)	7.11	0.68	15
EIII (both parents)	6.66	0.62	22

Homogenous vs. non-homogenous variance

OEHHA fit the continuous linear model to the individual dam data using the US Environmental Protection Agency's Benchmark Dose Software (BMDS). OEHHA reviewed the BMDS test statistics to determine if a homogenous or non-homogenous variance model should be used. For Staples, the homogenous variance model fit the data and was deemed appropriate (goodness-of-fit p= 0.91). In the case of Saillenfait, the homogenous variance model did not fit the data appropriately. The Sapphire Group (2009) chose to assume homogenous variance for the Saillenfait study. They dismissed the differences in variance as likely due to random chance, based on their conclusion that the variances fall within historical controls, and the variance of all treatments combined is "essentially the same as for control animals." However, the heterogeneity in the variance across exposure groups in the Saillenfait study is clearly apparent, particularly when the individual dam data are plotted (see OEHHA, 2009).

In BMDS, the only choice for modeling non-homogenous variance is a power function of the mean. As noted by The Sapphire Group (2009), this model does not fit the Saillenfait data appropriately, because the heterogeneity in the variance is not monotonic with exposure concentration. Applying the non-homogenous variance option in BMDS to model the Saillenfait data produces a goodness-of-fit test with a p-value of less than 0.1. Based on this test statistic, BMDS indicates, "You may want to consider a different variance model." OEHHA confirmed with the US Environmental Protection Agency (Gift, pers comm., 2009) that there is no alternative option to model the variance within BMDS. Neither variance modeling approach within BMDS fit appropriately. The BMCLs from both approaches are very similar: 75 ppm for non-homogenous variance and 77 ppm for homogenous variance.

If further analysis of the Saillenfait data is undertaken, OEHHA recommends that other variance models be examined to account for the observed non-monotonic heterogeneity in the variance. This could be done using software other than BMDS, such as R (available at: http://www.r-project.org/).

Control data for Saillenfait et al. (2003)

The Sapphire Group (2009) advocated using historical control data spanning from 1999 to 2007 to conduct the BMC analysis on the Saillenfait *et al*. (2003) study. They note that the mean from the Saillenfait controls for 1999-2007 is identical to the concurrent control mean from the 2003 study, and use of the historical control group should produce a more precise estimate. Concurrent control data are typically used in dose-response analyses, but OEHHA acknowledges that historical control information can provide additional insight. Controls from multiple developmental studies in the Saillenfait laboratory were pooled and reported by The Sapphire Group (2009). Using this historical control group, OEHHA derived BMCLs of 85 ppm (homogenous variance) and 81 ppm (non-homogenous variance) based on the grouped data. The grouped data were used because individual dam data were not obtained for the historical control group. As was true with the concurrent control group analysis, neither variance modeling approach within BMDS fit appropriately. The BMCLs derived using the concurrent control group were 77 ppm (homogenous variance) and 75 ppm (non-homogenous variance) based on the individual dam data. The BMCLs for the four cases were similar, ranging from 75 ppm to 85 ppm.

Choice of benchmark response (BMR)

For the NMP studies of body weight, OEHHA (2009) determined that a 5% relative deviation of the control mean is appropriate for the benchmark response (BMR). A 5% shift in the mean birth weight of a population is biologically significant. A shift in the mean indicates a shift in the population distribution (Kavlock *et al*., 1995), which means that some portion of the distribution would be pushed into the range of an adverse birth weight outcome. A 10% change in mean body weight is typically considered a marker for toxicity in adult animals. Shifts in body weights in fetuses or neonates are of even greater importance.

OEHHA acknowledges that there are many possible options for the BMR when performing a BMC analysis on continuous data. Other approaches discussed by the NMP Producers Group and The Sapphire Group *could* be appropriate for analyzing continuous data in certain cases,

including the choice of a BMR of 1 SD suggested in the EPA guidelines. However, the use of 1 SD as the BMR is a default approach and does not account for knowledge about the endpoint of concern. As noted by the NMP Producers Group (2009), the US EPA (2000) guidelines state that 1 SD is used as the BMR "**in the absence of any other idea** of what level of response to consider adverse" (emphasis added). In OEHHA's scientific judgment, there is indeed an "idea of what level of response to consider adverse" in the case of body weight effects and there is enough information to move away from the statistical default. The Sapphire Group (2009) quotes US EPA (2000) as stating that the BMCL should be reported using a 1 SD "for comparison purposes." OEHHA agrees that results based on a BMR of 1 SD, or some other fraction of the SD, could be reported for comparison purposes if desired by the risk assessor.

The Sapphire Group (2009) commented, "In the absence of a clear policy statement on this issue, many dose-response assessors adopted a BMR of 5% for continuous data based largely on published BMD work (Allen *et al*., 1994a, b; Kavlock *et al*., 1995). However, Allen *et al*. (1994a) makes the point of stating, 'Several examples and the discussion have focused on the 5% risk level; this stems from the similarity of the CBMD05 and the CNOAELs and the fact that 5% is in the middle of the range typically cited for BMD estimation (Crump, 1984; Kimmel and Gaylor, 1988). This should not be construed as a recommendation for that level of risk for BMD definition.'" The last part of the quote from Allen *et al.* (1994) was omitted by The Sapphire Group and is relevant for the discussion about BMR choice: "That decision also must be made in light of other choices required for the application of the BMD approach and will need to be considered together with other issues for establishing regulatory policy." The choice of 5% relative deviation as the BMR for analysis of body weight data took into account these other issues. As discussed above, OEHHA chose the BMR based on biological significance.

The Sapphire Group (2009) commented, "In this way, the 1 SD can be viewed as a data-derived BMR, while the use of a fixed value (*e.g.*, 5% or 10%) appears arbitrary and capricious." OEHHA agrees with The Sapphire Group that using 1 SD as the BMR is a data-driven approach. OEHHA does not agree that using a fixed value is "arbitrary and capricious." In fact, The Sapphire Group's decision to apply a statistical default in light of information indicating that a 5% relative change in mean body weight is biologically meaningful is itself an arbitrary choice. As noted by Slob (2002) "Instead of using the observed variation as a reference, the benchmark response can be defined in a biologically meaningful way as a particular change in the size of the effect that is considered acceptable or without adverse consequences for the subject. A toxicity study only allows for estimating the size of the effect (as a function of dose) as shown by the test animal *under the average of all experimental conditions* associated with the particular study.

From a biological point of view the most natural way of measuring an effect size is in terms of a percent change relative to the background value of the particular endpoint" (emphasis added by Slob). Foster (1995) states that "in order to define a BMD, an appropriate degree of change must be selected, for example a 5% reduction in mean foetal weight, based on the smallest measurable and toxicologically significant change." Foster also notes that "the BMDs are values that are **within the experimental range** of the study" (emphasis added by Foster *et al*., 1995). The Sapphire Group's BMR of 1 SD produced BMDs that were higher than the highest exposure concentrations (and the LOAELs) for both studies: 140 ppm (BMD) versus 120 ppm (highest exposure concentration and LOAEL) for Saillenfait, and 160 ppm versus 116 ppm for Staples.

The Sapphire Group (2009) commented that the results from the NMP studies argue for use of 1 SD as providing a "greater degree of concordance" in the BMCs. OEHHA agrees that the use of 1 SD as the BMR is well-suited for comparing between studies, and is recommended by US EPA for this purpose. However, in the current case, OEHHA's choice of the 5% relative deviation as the BMR for analyzing the NMP body weight studies rested on biological significance and not on a statistical argument. Further, concordance would not necessarily be expected between the Staples and Saillenfait studies. It is biologically plausible for the BMC from Staples to be lower than that for Saillenfait, as Staples had a longer exposure period and included exposure to both bucks and dams.

The NMP Producers Group (2009) commented, "The OEHHA Memorandum lists several publications to support its claim that a 5% BMR is biologically meaningful. However no such statement appears in any of the references provided by OEHHA." The NMP Producers Group further commented "Thus, CDPR equates a 5% BMR to a default and not to biological significance." These two comments are not correct. While it's true that the California Department of Pesticide Regulation (CDPR) identifies 5% relative deviation as a reasonable default BMR for continuous data analyses, CDPR explicitly discusses 5% relative deviation as a biologically meaningful approach for body weight studies: "A 5% BMR also appears reasonable for endpoints such as body weight changes since a 10% reduction is considered a marker of toxicity, an indication that the maximum tolerated dose (MTD) has been reached." OEHHA's citation of CDPR was specifically with regard to application of a 5% relative deviation as the BMR for the NMP body weight studies. OEHHA has not selected a default approach for analyzing continuous data. As noted in the OEHHA (2008) guidelines, it is OEHHA's policy that the BMR for continuous data is determined on a case-by-case basis using scientific judgment.

The NMP Producers Group (2009) commented, "Since none of the references discussed NMP, one is left to conclude that it is the policy of OEHHA to consider a 5% BMR as a biologically significant response for all chemicals." This is not correct for two reasons. First, for conducting benchmark dose analyses on continuous data, OEHHA's policy is as follows: "Other types of data, including continuous measures of toxic response, and data from epidemiological studies, require an appropriate benchmark response rate to be identified on a case by case basis" (OEHHA, 2008). Second, the 5% relative deviation was chosen as a biologically significant BMR for the NMP body weight studies and would not be considered a blanket approach for all types of endpoints associated with any chemical.

The NMP Producers Group (2009) stated, "Should the HEAC adopt this policy (*i.e.*, the use of a 5% relative deviation as the BMR), it will essentially eliminate the BMD tool from the PEL process since the outcome (*i.e.*, duplication of the study NOAEL) is predetermined." This is a misleading statement for multiple reasons. First, it is not true that the choice of a 5% relative deviation for analyzing continuous data gives a predetermined outcome of reproducing the NOAEL. In the current case, for example, the BMC analysis of the Saillenfait data results in a BMCL of 75 to 85 ppm (depending on the control group and the variance model) versus a NOAEL of 60 ppm, and for Staples results in a BMCL of 43 ppm versus a NOAEL of 50 ppm. The scientific literature has shown that a BMR of 5% relative deviation does not always reproduce the NOAEL. Second, while the goal is not to reproduce the NOAEL, it is certainly not appropriate for the BMCL to be associated with an adverse effect. Third, as OEHHA has noted multiple times in writing and verbally at HEAC meetings, the approach to analyzing continuous data is evaluated by OEHHA on a case-by-case basis. The choice of the 5% relative deviation as the BMR for NMP does not predetermine the choice of the BMR for other continuous data sets.

Data reported by dam are not interrelated

The NMP Producers Group (2009) notes that the BMD Guidance published by CDPR should not be applied to "nested data." As stated by CDPR (2004) nested data "are most commonly seen in reproductive and developmental endpoints when the response of fetuses from one litter are interrelated." The NMP Producers Group (2009) goes on to state, "These are the BMC endpoints at issue with NMP." It appears that the NMP Producers Group is implying that the CDPR (2004) document cannot be consulted for the NMP studies of reproductive and developmental endpoints. This is incorrect for multiple reasons. The CDPR document can be used as guidance for analyzing reproductive and developmental endpoints. The only issue is

how the data are reported. Specifically, if the data are reported by individual pup or fetus, then the data points are not independent and would be considered nested. If that were the case, then the correlation between pups or fetuses would need to be accounted for in the analysis. However, in the case of the NMP studies by Staples and Saillenfait, the data are reported as average litter body weight for each dam. By reporting the data in this way, all subjects (*i.e.*, individual dams) are independent and the data are not nested. The 5% relative deviation BMR choice recommended by CDPR (2004) is appropriate for body weight effects in independent subjects (dams in this case).

Choice of BMCL

OEHHA recommends using the BMCL from the Staples data, derived using a BMR of 5% relative deviation of the control mean. OEHHA (2003) identified Staples as the most appropriate and most sensitive study. Staples exposed both the dams and bucks for 12 weeks prior to meeting, continuing to expose the dams through gd 20. The Saillenfait study was limited to dam exposure during gd 6-20.

The NMP Producers Group (2008a) advocated for the use of Saillenfait citing the "higher NOAEL and superior quality (*e.g.*, the number of doses and animals on test)." The Group dismissed the exposure from gd 0-5 in Staples as being irrelevant because: (1) Exposure from gd 0-5 is not required according to "an EPA/OECD guideline" for a developmental toxicity study; (2) Exposure from gd 0-5 is not "relevant for a rapidly metabolized developmental toxin (*i.e.*, serum half-life of ~4 hours for the developmentally toxic form of NMP, the parent compound)"; and (3) Exposure from gd 0-5 is not "relevant for the toxic endpoint of concern (*i.e.*, fetal body weight – critical fetal period is gd 15-20)." OEHHA does not agree that Saillenfait is the superior study and has identified Staples as the most sensitive study of sufficient quality. Specific responses to The NMP Producers Group's (2008a) claims follow below.

First, the two NOAELs are not directly comparable, because the exposure parameters in the studies are substantially different. Exposure of both parents and a longer gestational exposure in the Staples study compared to the Saillenfait study might be anticipated to cause effects not detectable in the shorter study. Second, the two studies are similar in terms of group size. As noted by OEHHA (2009), "Both studies were sufficiently robust and large, with 4 dose groups containing 15 to 25 animals each, such that the probability of missing an effect was low (Slob *et al*., 2005)." Third, it is not the case that Saillenfait had a greater number of exposure groups – the two studies are equivalent in that regard. Fourth, the Staples study design goes beyond what

is required by the guidelines by including not only exposure from gd 0-5 but also by exposing both bucks and dams for 12 weeks prior to mating. The NMP Producers Group (2008a, 2008b; 2009) does not address the additional pre-mating exposure. The Staples experimental design in terms of exposure is clearly superior. Fifth, the half-life of a substance is not the only consideration in evaluating the relevance of a particular exposure scenario. The pharmacokinetic timetable does not necessarily match the timetable for an adverse effect. The empirical evidence from Staples suggests that exposure to both parents for 12 weeks prior to mating produced a stronger effect than exposure to each parent alone. The rapid clearance of NMP actually tends to confirm that this stronger effect is mediated by parental exposure and is in addition to any effect due to direct fetal exposure to NMP.

The NMP Producers Group (2008b) commented, "Saillenfait *et al.* used a greater number of animals per dose group; Staples used fewer. The greater uncertainty associated with fewer animals per dose will inevitably lead to a lower BMCL no matter the BMD approach used." This is not correct. OEHHA acknowledges that the numbers of litters per exposure group are not identical for Staples and Saillenfait. However, there are only a few more animals in Saillenfait groups. The difference between the two studies is 3-4 animals in the non-zero exposure groups (4, 4, and 3, for the low, mid and high exposure groups, respectively). The concurrent control group for Saillenfait is 24 animals while the concurrent controls in Staples is 39 animals. While there is a little more variability in the Staples study (SD is 9-10% of the mean for Staples, 5-8% for Saillenfait) the differences are small. To test the effect of animal number and variance, the Staples data (grouped) were analyzed in BMDS using the number of litters reported in Saillenfait (with the historical control group) and with standard deviations calculated to match Saillenfait (as a percentage of mean for each exposure group). The resulting BMCL was 45 ppm, which is essentially identical to the BMCL of 43 ppm produced with the actual Staples data. The primary difference between the two studies is the steeper slope of the dose-response curve in Staples compared to Saillenfait and not the numbers of animals per group.

The Sapphire Group (2009) suggests using the geometric mean to combine the results from the Staples (1990) and Saillenfait *et al*. (2003) studies. OEHHA has applied this method in the past (*e.g.*, cancer bioassays of equivalent quality) and is appropriate when there are two identical studies such that a distinction cannot be made. OEHHA acknowledges that there are a number of similarities between the two studies, and that the Saillenfait results provide support for the findings of Staples (OEHHA, 2003). However, Staples (1990) and Saillenfait *et al*. (2003) have very different exposure periods, as discussed above. Further, the measured endpoint for Staples (1990) and Saillenfait *et al.* (2003) is different. Staples measured pup weight at post-partum day

one, while Saillenfait measured fetal body weight at gd 21. It is not appropriate to combine the results of the Staples (1990) and Saillenfait *et al*. (2003) developmental toxicity studies using a geometric mean.

Conclusion

OEHHA has identified the Staples (1990) study as the most sensitive study of sufficient quality for the NMP benchmark concentration analysis. The Staples study design included pre-mating exposure to both parents and a longer gestational exposure period compared to Saillenfait *et al*. (2003). OEHHA has derived a BMCL for NMP of 43 ppm from the pup body weight data of Staples (1990). The Staples data are summarized below and the BMDS output is attached. This updated result for the Staples study takes into account two adjustments that were suggested by The Sapphire Group: use of measured concentrations and incorporation of a second concurrent control group.

Staples (1990)

Measured concentration (ppm)	***Pup bodyweight post-partum day one***		***Number of litters/dams***
	Mean (g)	***St. dev (g)***	
0 (CI and CII)	7.48	0.70	39
10.3	7.03	0.71	16
50.8	7.13	0.70	15
116 (EII only)	6.66	0.62	22

REFERENCES

Allen BC, Kavlock RJ, Kimmel CA, Faustman EM (1994). Dose-response assessment for developmental toxicity: II. Comparison of generic benchmark dose estimates with no observed adverse effect levels. *Fundam Appl Toxicol* **23**:487-495.

California Department of Pesticide Regulation (CDPR, 2004). *Guidance for Benchmark Dose (BMD) Approach – Continuous Data*. Medical Toxicology Branch, Department of Pesticide Regulation, California Environmental Protection Agency, Sacramento, CA. Available at: http://www.cdpr.ca.gov/docs/risk/bmdcont.pdf.

Foster PMD, Auton TR (1995). Application of benchmark dose risk assessment methodology to developmental toxicity: an industrial view. *Toxicol Lett* **82/83**:555-559.

Gift J (pers comm., 2009). Telephone call between Jeff Gift of the U.S. Environmental Protection Agency and Lindsey Roth of OEHHA. June 9, 2009.

Kavlock RJ, Allen BC, Faustman EM, Kimmel CA (1995). Dose-response assessment for developmental toxicity: IV. Benchmark doses for fetal weight changes. *Fundam Appl Toxicol* **26**:211-222.

NMP Producers Group (2008a). Comments on Permissible Exposure Limit Development for N-Methylpyrrolidone. Letter from Patrick Gibson of the NMP Producers Group to Bob Barish of the California Department of Occupational Safety and Health. April 25, 2008.

NMP Producers Group (2008b). NMP PEL: Comments on OEHHA presentation. Email from Ralph Parod of the NMP Producers Group to Bob Barish of the California Department of Occupational Safety and Health. October 1, 2008.

NMP Producers Group (2009). HEAC PEL Recommendation for N-Methylpyrrolidone. Letter from Ralph Parod of the NMP Producers Group to Bob Barish of the California Department of Occupational Safety and Health. May 5, 2009.

Office of Environmental Health Hazard Assessment (OEHHA, 2003). *Proposition 65 Maximum Allowable Dose Level (MADL) for Reproductive Toxicity for N-Methylpyrrolidone for dermal and inhalation exposures*. Available at: http://www.oehha.ca.gov/prop65/law/pdf_zip/NMPMADL2rev.pdf.

Office of Environmental Health Hazard Assessment (OEHHA, 2008). *Technical Support Document for the Derivation of Noncancer Reference Exposure Levels.* Available at: http://www.oehha.ca.gov/air/hot_spots/pdf/NoncancerTSDC061308.pdf.

Office of Environmental Health Hazard Assessment (OEHHA, 2009). N-Methylpyrrolidone Benchmark Dose Analysis. Memorandum from Sara Hoover of OEHHA to Dennis Shusterman of the California Department of Public Health from. March 5, 2009.

Poet T, Kirman CR, Gargas ML, Hinderliter PM (2009). Quantitative risk analysis for N-methylpyrrolidone using physiologically based pharmacokinetic and benchmark dose modeling. Draft submitted for publication.

The Sapphire Group (2009). Benchmark Concentration Calculations for N-Methyl-2-Pyrrolidone. Letter from Micheal Gargas of The Sapphire Group to Lisa Burchi of the NMP Producers Group. April 14, 2009.

Saillenfait AM, Gallissot F, Morel G (2003). Developmental toxicity of N-methyl-2-pyrrolidone in rats following inhalation exposure. *Food and Chemical Toxicology* **41**:583-588.

Slob W (2002). Dose-response modeling of continuous endpoints. *Toxicol Sci* **66**:298-312.

Slob W, Moerbeek M, Rauniomaa E, Piersma AH (2005). A statistical evaluation of the toxicity study designs for the estimation of the benchmark dose in continuous endpoints. *Toxicol Sci* **84**(1):167-185.

Staples R (1990). *1-Methyl-2-pyrrolidinone (NMP): Reproductive and developmental toxicity in the rat*. E.I. du Pont de Nemours and Company. Haskell Laboratory Report No. 294-90.

U.S. Environmental Protection Agency (U.S. EPA, 1991) *Guidelines for Developmental Toxicity Risk Assessment.* Risk Assessment Forum. Washington, DC, EPA/600/FR-91/001. Available at: http://www.epa.gov/NCEA/raf/pdfs/devtox.pdf.

U.S. Environmental Protection Agency (U.S. EPA, 2000). *Benchmark Dose Technical Guidance Document.* Risk Assessment Forum. Washington, DC, EPA/630/R-00/001. Available at: http://www.epa.gov/NCEA/pdfs/bmds/BMD-External_10_13_2000.pdf.

Attachment: BMDS (version 1.4.1b) Output

BMDS Output for Staples

```
BMDS MODEL RUN
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
   The form of the response function is:

   Y[dose] = beta_0 + beta_1*dose + beta_2*dose^2 + ...

   Dependent variable = RESPONSE
   Independent variable = Dose
   rho is set to 0
   Signs of the polynomial coefficients are not restricted
   A constant variance model is fit

   Total number of dose groups = 4
   Total number of records with missing values = 0
   Maximum number of iterations = 250
   Relative Function Convergence has been set to: 1e-008
   Parameter Convergence has been set to: 1e-008

                  Default Initial Parameter Values
                          alpha =     0.464224
                            rho =            0   Specified
                         beta_0 =      7.31839
                         beta_1 =  -0.00547457

           Asymptotic Correlation Matrix of Parameter Estimates
           ( *** The model parameter(s)  -rho have been estimated at a
boundary point, or have been specified by the user, and do not appear in the
correlation matrix )

                   alpha       beta_0       beta_1
     alpha             1    -1.8e-010    -1.2e-010
    beta_0    -1.8e-010            1        -0.63
    beta_1    -1.2e-010        -0.63            1

                                  Parameter Estimates
                                    95.0% Wald Confidence Interval
Variable   Estimate   Std. Err.   Lower Conf. Limit  Upper Conf. Limit
alpha      0.463938   .0684039          0.329869           0.598007
beta_0     7.38521     0.0909962        7.20686            7.56356
beta_1    -0.00622026  0.00150475      -0.00916951        -0.003271
```

Table of Data and Estimated Values of Interest

Dose	N	Obs Mean	Est Mean	Obs Std Dev	Est Std Dev	Scaled Res.
0	39	7.48	7.39	0.701	0.681	0.888
10.3	16	7.03	7.32	0.705	0.681	-1.74
50.8	15	7.13	7.07	0.695	0.681	0.365
116	22	6.66	6.66	0.616	0.681	-0.000142

Model Descriptions for likelihoods calculated

Model A1: Yij = Mu(i) + e(ij)
Var{e(ij)} = Sigma^2

Model A2: Yij = Mu(i) + e(ij)
Var{e(ij)} = Sigma(i)^2

Model A3: Yij = Mu(i) + e(ij)
Var{e(ij)} = Sigma^2
Model A3 uses any fixed variance parameters that
were specified by the user

Model R: Yi = Mu + e(i)
Var{e(i)} = Sigma^2

Likelihoods of Interest

Model	Log(likelihood)	# Param's	AIC
A1	-8.655329	5	27.310658
A2	-8.375428	8	32.750855
A3	-8.655329	5	27.310658
fitted	-10.671783	3	27.343566
R	-18.508588	2	41.017176

Explanation of Tests

Test 1: Do responses and/or variances differ among Dose levels?
(A2 vs. R)
Test 2: Are Variances Homogeneous? (A1 vs A2)
Test 3: Are variances adequately modeled? (A2 vs. A3)
Test 4: Does the Model for the Mean Fit? (A3 vs. fitted)
(Note: When rho=0 the results of Test 3 and Test 2 will be the same.)

Tests of Interest

Test	-2*log(Likelihood Ratio)	Test df	p-value
Test 1	20.2663	6	0.002483
Test 2	0.559803	3	0.9056
Test 3	0.559803	3	0.9056
Test 4	4.03291	2	0.1331

The p-value for Test 1 is less than .05. There appears to be a

```
difference between response and/or variances among the dose levels
It seems appropriate to model the data

The p-value for Test 2 is greater than .1.  A homogeneous variance
model appears to be appropriate here


The p-value for Test 3 is greater than .1.  The modeled variance appears
 to be appropriate here

The p-value for Test 4 is greater than .1.  The model chosen seems
to adequately describe the data

             Benchmark Dose Computation
Specified effect =           0.05
Risk Type        =     Relative risk
Confidence level =           0.95
             BMD =        59.3641
            BMDL =        42.8988
```