Department of Pesticide Regulation

Mary-Ann Warmerdam
Director

Arnold Schwarzenegger
Governor

M E M O R A N D U M

TO: Pamela Wofford
Senior Environmental Scientist
Environmental Monitoring Branch

FROM: Jing Tao, Ph.D. *Original signed by*
Environmental Scientist
Environmental Monitoring Branch
916-324-4201

DATE: November 9, 2009

SUBJECT: ANALYSIS OF SAMPLING INTENSITY FOR PROPOSED AIR MONITORING NETWORK BASED ON PESTICIDE MONITORING DATA FROM PARLIER, CALIFORNIA

INTRODUCTION

The Department of Pesticide Regulation (DPR) plans to set up an air monitoring network to regularly monitor multiple pesticides in several communities over the next five or more years. DPR will use the data gathered to evaluate and improve protective measures against pesticide exposure. The monitoring would consist of two to four sites in each community and each site would collect 24-hour samples on a recurring schedule. Several issues arise concerning the frequency of sampling which directly affects the number of samples and the personnel costs and hence, the number of different communities which can be sampled. These concerns include: (a) whether it is necessary to sample in more than one location within each proposed community; (b) whether it is necessary to sample on weekends or particular days of the week; and (c) if sampling can be conducted less than once-per-week. In order to design an efficient and effective monitoring protocol, we analyzed data from a recent DPR monitoring study in Parlier, California.

The Parlier study monitored 31 pesticides and pesticide breakdown products in 3 ambient locations from January 3 to December 26, 2006. Samples were collected for 24 hours three consecutive days a week at each of the three sites. The weekly starting day varied each week. With regard to the proposed air monitoring scheme, we asked the following questions of the Parlier monitoring data:

- Question 1: Were samples from the three Parlier locations different from each other?
- Question 2: Were samples on weekend days different from other days of the week?
- Question 3: Were samples on different days of the week different from each other?
- Question 4: Would sampling every week differ from sampling every other week?

The analysis described in this memorandum provides answers for these questions.

METHODS

Data description

The Parlier study used in this analysis monitored 31 pesticides and pesticide breakdown products based on their: (a) toxicity, (b) vapor pressure (volatility), (c) use, (d) availability of sampling and laboratory methods, and (e) ability to be included in a multiple analysis method. The results showed that 17 chemicals were not detected at a level above the method detection limit ([MDL], the smallest amount of the chemical that can be identified in a sample with the employed method). There nondetected concentrations were referred as "nd." Six chemicals were detected as "trace," which was a level between the MDL and the limit of quantitation ([LOQ], the smallest amount of a chemical that can be measured). Eight chemicals were detected at quantifiable levels. Three pesticides with the highest percentages of concentrations above the MDL were used in this analysis: methyl isothiocyanate (MITC) with 84 percent (%) detect concentrations, Chlorpyrifos with 64%, and Diazinon with 32%.

We adjusted the original dataset in two different ways. The first adjusted dataset contained all of the original concentration values but replaced "nd" with the average MDL and "trace" with the midpoint between MDL and LOQ. The second dataset consisted of only two values: 1 if the original value was quantifiable or "trace" concentrations and 0 if the original value was "nd." This second adjustment helps to determine the probability of detecting chemicals with different sampling frequency instead of the concentration distribution.

The normality was tested on the first dataset with procedure of CAPABILITY by SAS 9.1. This procedure consists of four methods: Kolmogorov-Smirnov, Cramer-von Mises, Anderson-Darling and Chi-Square. Small p-values suggest non-normal distributions.

Question 1: Difference between sampling stations

Air-sampling stations were set up at three elementary schools in Parlier: John C. Martinez (MAR, northwest part of town), S. Ben Benavidez (BEN, central), and Cesar E. Chavez (CHA, southeast). To answer this question, three analyses were conducted:

<u>Method 1: correlation coefficient</u>
The correlation coefficients were calculated on chemical concentrations for three pairs of sampling stations. High correlation would suggest the consistency of sampling in two stations. Otherwise, difference may exist between them.

<u>Method 2: nonparametric statistical method (Kruskal-Wallis Method)</u>
Reference: Hollander, M. & Wolfe, D.A., Nonparametric Statistical Methods, Second Edition (1999), Wiley.

Kruskal-Wallis method is a nonparametric multiple comparison procedure based on pairwise ranking. It was applied to compare the significant difference between three sampling sites for each of three pesticides using Minitab 15. The first dataset was used.

The hypothesis of the procedure:
$H_0 : \tau_1 = \tau_2 = \tau_3$, where τ is the effect of three sampling sites
$H_1 : \tau_1, \tau_2, \tau_3$ not all equal

Kruskal-Wallis Statistic

$$H = \left(\frac{12}{N(N+1)} \sum_{j=1}^{k} \frac{R_j^2}{n_j} \right) - 3(N+1)$$

Where N is the total number of observations; n_j is the observation number in j sampling site; $R_j = \sum_{i=1}^{n_j} r_{ij}$, let r_{ij} denote the rank of observation X_{ij} in the joint ranking.

Reject H_0 if $H \geq h_\alpha$, where h_α is chosen to make the type I error probability equal to $\alpha = 0.05$.

Method 3: comparison of proportion of detects
The second dataset were used in the method 3. The proportions of detect values (1) were calculated for each location and compared to answer if chemicals can be detected with the same probability in different location.

Question 2: Difference between weekend and other days of the week

The second dataset was analyzed to answer if different sampling frequency impacts the detection probability. For this question, the data were divided into two categories: weekend (1) and other days of week (0) and summarized in contingency tables (Table 1). The proportion of each cell was calculated and compared for their significant difference.

Statistical tests were also conducted. Since the counts of both weekend and weekday are large (>30), the proportions of positive values for weekend and weekday have approximately normal distribution. Therefore, the difference between them is normal and z value can be calculated as:

$$z = \frac{\hat{p}_0 - \hat{p}_1}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_0}+\frac{1}{n_1})}}$$ For example: in Chlorpyrifos

$\hat{p}_0 = \frac{277}{432} = 0.64$, $\hat{p}_1 = \frac{22}{36} = 0.61$, $\hat{p} = \frac{299}{468} = 0.64$, $n_0 = 432$, $n_1 = 36$

The statistic z is compared with the standard normal distribution table.

Table 1. Counts of detect (1) and nondetect (0) sampled on weekends (1) and other days (0).

Chlopyrifos	Weekend		
Count	0	1	Total
0	155	14	169
1	277	22	299
Total	432	36	468

Diazinon	Weekend		
Count	0	1	Total
0	296	23	319
1	136	13	149
Total	432	36	468

MITC	Weekend		
Count	0	1	Total
0	72	4	76
1	360	32	392
Total	432	36	468

Question 3: Difference between days of the week

For this question, the second dataset was grouped by each sampling start day within a week. Monday was 1, Tuesday 2, . . ., and Sunday 7 (Table 2). No sampling started on Saturday. The proportions of positive values were calculated and compared.

Statistical testing was conducted to address any significant difference of a day from others within a week. Let the observed estimate O_{ij} = count {sample value from jth day fall in ith category (detect or nondetect)}. Then the statistic: $\sum_i \sum_j \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$ has approximately χ^2_{df} distribution. Where E_{ij} is expected count and equal to $\frac{n_{i.} n_{.j}}{n_{..}}$, df = (I-1)(J-1) = 5.

Table 2. Counts of detect (1) and nondetect (0) sampled on each day of the week.

Chlorpyrifos	Day of the week						
Counts	1	2	3	4	5	7	Total
0	32	50	38	24	11	14	169
1	49	70	85	51	22	22	299
Total	81	120	123	75	33	36	468

Diazinon	Day of the week						
Counts	1	2	3	4	5	7	Total
0	52	86	84	52	22	23	319
1	29	34	39	23	11	13	149
Total	81	120	123	75	33	36	468

MITC	Day of the week						
Counts	1	2	3	4	5	7	Total
0	12	22	26	9	3	4	76
1	69	98	97	66	30	32	392
Total	81	120	123	75	33	36	468

Question 4: Difference between sampling every week and every other week

The second dataset was categorized as odd week and even week and the counts of detects and nondetects are listed in Table 3. The proportions of detect values were calculated and compared.

Table 3. Counts of detect (1) and nondetect (0) sampled in odd and even weeks.

Chlorpyrifos	Week		
Counts	even	odd	Total
0	84	85	169
1	150	149	299
Total	234	234	468

Diazinon	Week		
Counts	even	odd	Total
0	163	156	319
1	71	78	149
Total	234	234	468

MITC	Week		
Counts	even	odd	Total
0	38	38	76
1	196	196	392
Total	234	234	468

RESULTS AND DISCUSSION

Data distribution

The result of normality test is shown in Table 4. All the p-Values < 0.01, which indicated that the monitoring data were not normally distributed. In addition, 90% of Chlorpyrifos data, 91% of Diazinon, and 70% of MITC were assigned numbers ("nd" or "trace") in this test and this feature probably contributed to the nonnormality. Therefore, the statistical methods based on numerical concentrations with assumption of normality can not be directly applied on the measured data in the Parlier study. Correlation coefficients, contingency tables and nonparametric methods were conducted in this work.

Table 4. Normality test on chemical concentrations of three pesticides.

Chlorpyrifos: Goodness-of-Fit Tests for Normal Distribution					
Test	Statistic		DF	p-value	
Kolmogorov-Smirnov	D	0.353		Pr > D	<0.010
Cramer-von Mises	W-Sq	9.724		Pr > W-Sq	<0.005
Anderson-Darling	A-Sq	51.490		Pr > A-Sq	<0.005
Chi-Square	Chi-Sq	12035.821	10	Pr > Chi-Sq	<0.001

Diazinon: Goodness-of-Fit Tests for Normal Distribution					
Test	Statistic		DF	p-value	
Kolmogorov-Smirnov	D	0.403		Pr > D	<0.010
Cramer-von Mises	W-Sq	23.582		Pr > W-Sq	<0.005
Anderson-Darling	A-Sq	113.975		Pr > A-Sq	<0.005
Chi-Square	Chi-Sq	1945.073	10	Pr > Chi-Sq	<0.001

MITC: Goodness-of-Fit Tests for Normal Distribution					
Test	Statistic		DF	p-value	
Kolmogorov-Smirnov	D	0.441		Pr > D	<0.010
Cramer-von Mises	W-Sq	31.015		Pr > W-Sq	<0.005
Anderson-Darling	A-Sq	147.105		Pr > A-Sq	<0.005
Chi-Square	Chi-Sq	128.737	10	Pr > Chi-Sq	<0.001

Question 1: Difference between sampling stations

Result 1: correlation coefficient
The correlation coefficients were calculated for daily measurement, weekly average, and monthly average data (Table 5). The data from different sites were highly related with each other for pesticide Chlorpyrifos, and moderately for Diazinon. The weekly and monthly average concentrations showed increased correlation between three sites for these two pesticides. This indicated that different sampling sites followed the similar pattern on weekly, and monthly exposure, which may related with the chronicle pattern of pesticide use.

Table 5. Correlation between pesticide concentrations of three sampling stations: Benavidez (BEN), Chavez (CHA) and Martinez (MAR).

Daily data

Correlation Coefficient	Pesticide		
	Chlorpyrifos	*Diazinon*	*MITC*
BEN v.s. CHA	0.76	0.50	0.49
BEN v.s. MAR	0.80	0.63	0.75
CHA v.s. MAR	0.75	0.77	0.79

Weekly average

Correlation Coefficient	Pesticide		
	Chlorpyrifos	*Diazinon*	*MITC*
BEN v.s. CHA	0.82	0.52	0.47
BEN v.s. MAR	0.91	0.81	0.70
CHA v.s. MAR	0.83	0.79	0.88

Monthly average

Correlation Coefficient	Pesticide		
	Chlorpyrifos	*Diazinon*	*MITC*
BEN v.s. CHA	0.95	0.67	0.18
BEN v.s. MAR	0.96	0.87	0.51
CHA v.s. MAR	0.94	0.88	0.84

The coefficient between BEN and CHA was 0.49 in MITC daily data and dropped to 0.18 for MITC monthly average data. The decreasing coefficient was also found between BEN and MAR for MITC in daily versus monthly correlation. This result came from an extreme MITC concentration of 5012.4 mg/L sampled at BEN in May. This value was seven times that of the second highest concentration of MITC (723.6 mg/L at BEN in May) and nine times the third (548.48 mg/L at MAR in May). Consequently, the average concentration of May was 391.48 in BEN, much higher than 26.36 in CHA and 48.81 in MAR. Since other monthly averages tracked more closely to each other, large magnitude differences in one month caused the statistically weak correlation. Excluding May, the correlation coefficient was 0.89 between BEN and CHA, 0.97 BEN and Mar, and 0.89 CHA and MAR for monthly data. Therefore, aside from this single extreme concentration, three sampling sites showed reasonably high correlation to each other.

Result 2: nonparametric method

Nonparametric method does not require the traditional assumption that the underlying population of the data is normal. Median, instead of mean, is used as statistical estimate and to compare different treatment group/effect. Therefore, it is applicable in Parlier study since the data showed non-normal distribution.

The result of Kruskal-Wallis method is shown in Table 6. The large p-values indicated that no significant difference between sampling locations according to nonparametric method.

Table 6. Result of nonparametric Kruskal-Wallis test for comparison of three sampling stations.

Pesticides	Statistic	
	H	p-value
Chlorpyrifos	0.22	0.90
Diazinon	0.04	0.98
MITC	1.15	0.56

Result 3: comparison of proportion of detects

Table 7 lists the proportion of detect values of three pesticides. The result shows that three locations had the same probability to positively detect samples with close proportions of positive values.

Table 7. Proportions of detects in three sampling stations.

Sampling Stations	Proportion		
	Chlorpyrifos	*Diazinon*	*MITC*
Martinez	0.66	0.33	0.84
Benavidez	0.62	0.32	0.85
Chavez	0.64	0.31	0.82
Average	0.64	0.32	0.84

Question 2: Difference between weekend and other days of the week.

The proportions of positive values were close to each other for weekend and weekday (Table 8). The results of statistic z were also listed in the table. The absolute values of z were very small compared to 1.65 at $\alpha = 0.10$ and suggested no statistically significant difference.

Table 8. Statistical analysis on proportions of detect (1) and nondetect (0) sampled on weekends (1) and other days of the week (0).

Chlorpyrifos	Weekend		$\hat{p}$	z
Proportion	0	1		
0	0.36	0.39	0.36	-0.36
1	0.64	0.61	0.64	0.36
Total	1.00	1.00		

Diazinon	Weekend		$\hat{p}$	z
Proportion	0	1		
0	0.69	0.64	0.68	0.57
1	0.31	0.36	0.32	-0.57
Total	1.00	1.00		

MITC	Weekend		$\hat{p}$	z
Proportion	0	1		
0	0.17	0.11	0.16	0.87
1	0.83	0.89	0.84	-0.87
Total	1.00	1.00		

Question 3: Difference between days of the week

For three pesticides, proportions of positive values are listed and they are close over a week (Table 9). The statistical test estimates χ^2 are 4.24 for Chlorpyrifos, 1.65 for Diazinon, and 5.60 for MITC. These values are very small compared to 9.24 at $\alpha = 0.10$ and indicate no significant difference between each day within a week.

Table 9. Proportions of detects sampled on different days of the week.

Proportion	Pesticide		
Day of the week	*Chlorpyrifos*	*Diazinon*	*MITC*
Mon	0.60	0.36	0.85
Tue	0.58	0.28	0.82
Wed	0.69	0.32	0.79
Thr	0.68	0.31	0.88
Fri	0.67	0.33	0.91
Sun	0.61	0.36	0.89
Average	0.64	0.32	0.84

Question 4: Difference between sampling every week and every other week

Table 10 shows that the detect values appeared with the same proportion in the even weeks and odd weeks. Therefore, there was no difference between sampling in even and odd weeks, and also between sampling every week and every other week.

Table 10. Proportions of detects sampled in odd and even weeks.

Proportion	Pesticide		
Week	*Chlorpyrifos*	*Diazinon*	*MITC*
Odd	0.64	0.30	0.84
Even	0.64	0.33	0.84
Average	0.64	0.32	0.84

CONCLUSION

The monitoring data of three pesticides in the Parlier study were analyzed to suggest an appropriate sampling set-up for the air monitoring network. According to the statistical analysis result, different days of week had the same probability to obtain samples with detectable concentrations. Sampling every other week was enough to detect the same proportion of positive values with sampling every week. Although one of the nine possible pairwise correlations between three sites showed fair relationship, further analysis and nonparametric method showed that samples from three sites were not significantly different with each other. They were also consistent in the probability to sample positive values.