

# **Comparison of Methods for the Determination of Alkyl Phosphates in Urine**

# **Comparison of Methods for the Determination of Alkyl Phosphates in Urine**

by

Ryan R. James  
Atmospheric Science and Applied Technology  
Battelle Memorial Institute  
Columbus, Ohio 43201

Stephen C. Hern, Gary L. Robertson, Brian A. Schumacher  
National Exposure Research Laboratory  
Las Vegas, NV 89193

EPA Contract 68-D-99-011

EPA Project Officer

Ellen W. Streib  
National Exposure Research Laboratory (MD-56)  
Research Triangle Park, North Carolina 27711

National Exposure Research Laboratory  
Office of Research and Development  
U.S. Environmental Protection Agency  
Research Triangle Park, North Carolina 27711

## **Notice**

The information in this document has been funded wholly or in part by the United States Environmental Protection Agency under Contract 68-D-99-011 to Battelle Memorial Institute. It has been subjected to the Agency's peer and administrative review and has been approved for publication as an EPA document. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

## Foreword

The mission of the National Exposure Research Laboratory (NERL) is to provide scientific understanding, information, and assessment tools that will quantify and reduce the uncertainty in EPA's exposure and risk assessments for environmental stressors. These stressors include chemicals, biologicals, radiation, and changes in climate, land use, and water use. The Laboratory's primary function is to measure, characterize, and predict human and ecological exposure to pollutants. Exposure assessments are integral elements in the risk assessment process used to identify populations and ecological resources at risk. The EPA relies increasingly on the results of quantitative risk assessments to support regulations, particularly of chemicals in the environment. In addition, decisions on research priorities are influenced increasingly by comparative risk assessment analysis. The utility of the risk-based approach, however, depends on accurate exposure information. Thus, the mission of NERL is to enhance the Agency's capability for evaluating exposure of both humans and ecosystems from a holistic perspective.

The National Exposure Research Laboratory focuses on four major research areas: predictive exposure modeling, exposure assessment, monitoring methods, and environmental characterization. Underlying the entire research and technical support program of the NERL is its continuing development of state-of-the-art modeling, monitoring, and quality assurance methods to assure the conduct of defensible exposure assessments with known certainty. The research program supports its traditional clients – Regional Offices, Regulatory Program Offices, ORD Offices, and Research Committees – and ORD's Core Research Program in the areas of health risk assessment, ecological risk assessment, and risk reduction.

Gary J. Foley  
Director  
National Exposure Research Laboratory

## Abstract

Organophosphorous (OP) pesticides have been used heavily in the United States and have been detected in dust, handwipes, drinking water, food, and air indicating human exposure pathways. Once inside the body, these pesticides are metabolized mostly to one of six alkyl phosphate compounds: dimethylphosphate, dimethylphosphorodithioate, dimethylphosphorothioate, diethylphosphate, diethylphosphorothioate, and diethylphosphorodithioate. These metabolites have been detected in urine and the quantity of these metabolites in urine has been shown to correlate with the level of pesticide dose that a person has experienced. Therefore, the measurement of these urinary metabolites can be used to assess and compare exposure. Unfortunately, this measurement is not straightforward. To characterize the performance of four existing analytical methods used to analyze urine samples for the six urinary alkyl phosphate metabolites of OP pesticides, an interlaboratory comparison study was done.

Thirty-five urine samples fortified with various concentrations of the alkyl phosphate metabolites were distributed to four laboratories that have developed and implemented analytical methods to measure these compounds. The results provided by each laboratory were analyzed by an analysis of variance (ANOVA) model to satisfy two objectives. The first was to identify those compounds where statistically significant differences existed (at the 0.05 level) in the reported measurements between the concentration levels within each laboratory, in order to determine the approximate detection threshold of each laboratory. The second was to determine when statistically significant differences existed in the reported measurements between the analytical methods, in order to compare the overall performance of the participating laboratories and hence, the different methods.

The study resulted in the following recommendations regarding urinary alkyl phosphate analyses:

- Given the variability of the data, especially at low concentrations, care should be used in interpreting relatively small differences between samples.
- Although there is considerable within and between laboratory variability, all of the laboratories could distinguish between samples containing low, medium, and high levels of alkyl phosphate metabolites.
- Given the sample to sample variability, especially among the blind replicates, preparing and analyzing each sample in duplicate will improve data quality.
- It is recommended that a performance evaluation sample of known concentration be developed and analyzed with each batch of samples to provide assurance the method is performing as expected.

The work reported herein was performed by Battelle Memorial Institute under U.S. Environmental Protection Agency Contract 68-D-99-011. Work was completed as of May 15, 2003.

## Contents

Foreword .....	iii
Abstract .....	iv
Tables .....	vi
Acknowledgments .....	vii
Chapter 1 Introduction .....	1
Chapter 2 Conclusions .....	3
Chapter 3 Experimental Methods .....	5
Laboratory Participation .....	5
Materials and Sample Handling .....	5
Experimental Design .....	6
Chapter 4 Results and Discussion .....	10
Statistical Differences Between Concentration Levels .....	10
Accuracy of Reported Concentrations .....	13
Statistical Differences Between the Laboratories .....	15
References .....	19
Appendix A Descriptions of the Participating Laboratories' Methods for Measuring Alkyl Phosphates in Urine .....	A-1
Appendix B Statistical Methods and Results .....	B-1
Appendix C Raw Data .....	C-1

## Tables

1	Common urinary alkyl phosphates .....	1
2	Concentrations corresponding to each mix and spiking level .....	7
3	Detection limits reported by each participating laboratory .....	7
4	Sample testing matrix for each participating laboratory, according to the spiking concentration levels associated with each compound mix .....	8
5	The randomized order of sample testing specified for each laboratory .....	9
6	Summary of concentration level effects for each lab and for each alkyl phosphate target compound .....	11
7	Lowest spiking level of alkyl phosphate target analytes that were significantly larger than the unspiked level .....	13
8	Recoveries of alkyl phosphate target analytes in spiked urine samples .....	14
9	Range of and average recoveries across all target analytes and all detectable concentration levels .....	15
10	Lowest spiked concentration of alkyl phosphate target analytes that were significantly larger than the unspiked level, the average percent recovery at that spiking level in parentheses, and the reported detection limit for each participating laboratory .....	15
11	Summary of lab effects at low #2, medium, and high concentration levels .....	17

## **Acknowledgments**

We thank Pacific Toxicology Laboratories (Woodland Hills, CA), the University of Washington Department of Environmental Health (Seattle, WA), Centers for Disease Control and Prevention (Atlanta, GA), and Centre de Toxicologie Institut National de Sante Publique du Quebec (Sainte-Foi, Quebec, Canada) for participating in this study. Major contributions to the research effort were also made by Battelle staff members - Robert Lordo, Zhenxu (James) Ma, Donald Kenny, and Julie Sowry.



# Chapter 1

## Introduction

Organophosphorus (OP) insecticides are among the most widely used and frequently detected pesticides in the U.S. (Lewis et al., 1988; Fortmann et al., 1991; Murphy et al., 1983). They have been detected in dust, handwipes, drinking water, food, and air indicating human exposure pathways. Upon entering the body, most organophosphorus pesticides are metabolized to yield one or more of the six common alkyl phosphates shown in Table 1. These metabolites have been detected in urine and the quantity of these alkyl phosphate metabolites excreted in human urine has been shown to provide a measure of pesticide dose (Morgan et al., 1977; Franklin et al., 1981; Bradway et al., 1977). Therefore, the measurement of these urinary metabolites can be used to assess and compare exposure. Unfortunately, this measurement is not straightforward-hence, this study was undertaken to evaluate the existing analytical methods. This study characterized the performance of four different laboratory methods, described in Appendix A, that were used to analyze urine samples for the six urinary alkyl phosphate metabolites of OP pesticides. This report presents the methods and results of this study. For this study, we recruited laboratories that had each developed and implemented

**Table 1. Common urinary alkyl phosphates**

Name	Acronym
Dimethylphosphate	DMP
Dimethylphosphorodithioate	DMDTP
Dimethylphosphorothioate	DMTP
Diethylphosphate	DEP
Diethylphosphorothioate	DETP
Diethylphosphorodithioate	DEDTP

a specific analytical method to analyze urine samples for each of the above six alkyl phosphates. These samples were fortified with the six alkyl phosphate compounds at concentration levels unknown to the participating laboratories. The measurement data generated by each laboratory were analyzed in order to make statistical comparisons of the results across analysis methods for each of the six alkyl phosphates.

Chapter 2 summarizes the results of this method comparison study for each participating laboratory and presents an overview of their performance compared with one another. The experimental design of the study is described in Chapter 3. This includes the amounts of alkyl phosphate metabolites added to the urine samples and the experimental matrix of urine samples containing various concentrations of the target analytes. Chapter 4 explains the results of the statistical analysis of the concentration data submitted by each participating laboratory. The focus is on determining significant differences between the concentration levels within each laboratory and significant differences between the performance of each laboratory at each concentration level. The ability of each laboratory to measure the target analytes near their reported detection limits is also

discussed. Appendix A provides a summary of the analytical methods used by each laboratory. Appendix B is the complete report of the statistical analysis of the data, which is summarized and discussed in Chapter 4. Appendix C contains the analytical results data provided by each participating laboratory.

## **Chapter 2**

### **Conclusions and Recommendations**

#### **Overall conclusions and recommendations include:**

- 1) Given the variability of the data, especially at low concentrations, care should be used in interpreting relatively small differences between samples.
- 2) Although there is considerable within and between laboratory variability, all of the laboratories could distinguish between samples containing low, medium, and high levels of alkyl phosphate metabolites.
- 3) The DMP results were problematic for all of the laboratories. This may have been related to the preparation and handling of the samples prior to shipment to the laboratories rather than a problem with the analytical methods.
- 4) Given the sample to sample variability, especially among the blind replicates, preparing and analyzing each sample in duplicate will improve data quality.
- 5) It is recommended that a performance evaluation sample of known concentration be developed and analyzed with each batch of samples to provide assurance the method is performing as expected.

#### **Specific conclusions from the statistical analysis**

- 1) Lab A reported concentrations that were significantly different from (and higher than) the unspiked samples for DMDTP, DEP, and DEDTP at the Low #1 and #2 concentration levels, DETP at the Low #2 concentration level, and for all of the target analytes except DMP at the Medium and High concentration levels. The overall average recovery for Lab A was 103% with a standard deviation of 39%.
- 2) Lab B reported concentrations that were significantly different from (and higher than) the unspiked samples for all the target analytes except DMP at the Medium and High concentration levels. However, with the exception of DEP and the High level concentrations of DMTP and DETP, Lab B's recoveries were generally much greater than 100%.
- 3) Lab C reported concentrations that were significantly different from the unspiked samples for DEDTP, DMTP, and DETP at the Low #2 concentration level and for all the target analytes except DMP at the Medium and High concentration levels. The overall average recovery for Lab C was 88% with a standard deviation of 25%.
- 4) Lab D reported concentrations that were significantly different from the unspiked samples for DMTP and DETP at the Low #2 concentration level, and for all of the target analytes except DMP at the Medium and High concentration levels. However, for DMDTP and DEP, the High level concentrations were not statistically different from the Medium level. The overall average recovery for Lab D was 100% with a standard deviation of 62%.

- 5) Lab A reported concentrations of spiked samples near its detection limits that were significantly different those reported for the unspiked samples for all the target analytes except for DMTP. For Lab B, only those reported concentrations well above its reported detection limits were found to be significantly different from those reported for the unspiked samples for all the target analytes. Labs C and D reported concentrations of spiked samples near their detection limits that were significantly different from those reported for the unspiked samples for all the target analytes (Lab D does not measure DEDTP) except for DMDTP and DEP.
- 6) None of the laboratories at any spiking level reported concentrations for DMP that were significantly different from the concentrations determined in the unspiked urine samples. It is unclear why the results for DMP were poor across all four laboratories. A review of the solution preparation records confirmed the addition of DMP to the spiked urine. Additional study is recommended to investigate the occurrence of a possible matrix interference in the urine that may keep the DMP from being extracted or that causes degradation of DMP in the urine. Also, verification of the purity of the DMP standard used to make the urine solutions needs to be done to investigate the possibility that an impure standard was the reason for the poor results for DMP.

## **Chapter 3**

### **Experimental Methods**

#### **Laboratory Participation**

Four laboratories agreed to participate in the study, namely Pacific Toxicology Laboratories (Woodland Hills, CA), the University of Washington Department of Environmental Health (Seattle, WA), Centers for Disease Control and Prevention (Atlanta, GA), and Centre de Toxicologie Institut National de Sante Publique du Quebec (Sainte-Foi, Quebec, Canada). In no particular order, they are identified as Labs A through D in this report. Each of the participating laboratories submitted information on detection limits, details of their method, required sample size, and costs to participate. Appendix A provides a description of each laboratory's method for measuring alkyl phosphates in urine.

#### **Materials and Sample Handling**

Four of the target compounds were available from commercial vendors, DMP (Pfaltz and Bauering, Waterbury, CT) DEP (Chem Service, West Chester, PA), and DETP and DEDTP (Aldrich Chemical, Milwaukee, WI); and the remaining two test compounds, DMTP and DMDTP, were obtained from Applichem GmbH, Germany as a custom synthesis.

All of the test solutions were prepared either from a commercially available pooled urine (American Biological Technologies, Seguin, TX) or a synthetic urine (formulation obtained from Centers for Disease Control and Prevention). Samples prepared in the pooled urine were fortified at only the Medium and High concentration level in order to avoid interference from background levels of the alkyl phosphate metabolites in the pooled urine. Samples prepared in the synthetic urine were fortified with all concentration levels. Unspiked samples were prepared from both the pooled and synthetic urine, but only those prepared in the synthetic urine were included in the statistical analysis so significant differences between the unspiked levels and the lower concentration levels could be evaluated.

Stock solutions of the alkyl phosphate metabolites were prepared by weighing 10 to 15 mg of the solid compounds into a weighing boat using an analytical balance (Mettler AE1660). The calibration of the balance was confirmed with 5 and 100 mg standard weights prior to use and the exact weights of each target analyte was recorded in a laboratory notebook to the nearest tenth of a milligram. The mixture was dissolved in distilled water and two tenfold dilutions were performed to prepare working stock solutions. Appropriate volumes of the stock solutions were pipetted into each sample using an Eppendorf pipette. To assure homogeneity of the samples between the laboratories, each sample was prepared from a single volume of urine and allocated into the sample containers for each of the participating laboratories. Unique identifier numbers were assigned to each sample.

After preparation, all samples were stored in a freezer at -20°C to protect against degradation of the alkyl phosphate compounds. Each set of samples was shipped under dry ice for next-day delivery to the participating laboratories to ensure that the samples remained frozen during shipment.

Special care was taken to protect against breakage and to conform with all state and federal regulations for transport of biohazardous material. All of the participating laboratories were required to store the urine samples in a -20°C freezer prior to analysis. All of the laboratories were contacted prior to shipment of samples so receipt of the samples on dry ice was ensured. Lab A received three sets of 35 samples for replicate analysis while Labs B through D received a single set.

### **Experimental Design**

The six alkyl phosphate target compounds were prepared in two spiking mixtures. The compounds in each mix were as follows:

- Mix A: DMP, DMDTP, DEP, DEDTP
- Mix B: DMTP, DETP

When spiking the urine samples with a given mix, all compounds within that mix were represented at the same concentration level. Table 2 shows the four concentration levels of each mix, plus an unspiked level, that were determined to be sufficient to characterize method performance for each lab (and, equivalently, each analytical method) over a range of concentration levels for a given compound and in the company of compounds from the other mix at various concentration levels. Thus, there were  $5 \times 5 = 25$  different types of samples prepared in this study, corresponding to each combination of the two mixes at the following five concentration levels:

- Unspiked (authentic pooled and synthetic urine)
- Low #1 (spiked near the detection limit for Lab A; Table 3)
- Low #2 (spiked near detection limit for the other three labs)
- Medium (spiked at approximately two to five times the highest detection limit reported by the participating laboratories)
- High (spiked at 200 µg/L).

Due to a laboratory error during sample preparation, the samples that were supposed to be spiked with Mix B at the Low #1 concentration level were instead spiked at the Low #2 level. Thus, there were twice as many samples spiked at the Low #2 concentration level for Mix B than at the unspiked, medium, and high levels, and no sample was spiked at the Low #1 level for Mix B. Mix A compounds were spiked at their specified level for each sample, and thus were unaffected by this laboratory error.

**Table 2. Concentrations corresponding to each mix and spiking level (µg/L)**

Compound		Low Level #1	Low #2 Level	Medium Level	HighLevel
Mix A	DMP	1.00	2.00	50.0	200
	DMDTP	1.04	2.08	52.0	208
	DEP	1.40	2.80	70.0	280

	<b>DEDTP</b>	1.35	2.70	67.5	270
<b>Mix B</b>	<b>DMTP</b>	NA <sup>a</sup>	3.82	23.9	191
	<b>DETP</b>	NA	4.44	27.8	222

<sup>a</sup> There was only one low level of Mix B.

Table 3 presents the reported detection limits for each compound for the four participating labs. Note that because the Low #1 spike level was considerably below the detection limits of all but Lab A, samples spiked with Mix A at this level were expected to resemble unspiked samples (with regard to Mix A) for those labs with higher detection limits. Meanwhile, for Lab A, analysis of the Mix A compounds at the Low #1 level was expected to provide information on performance at a level close to their detection limit.

**Table 3. Detection limits reported by each participating laboratory (µg/L)**

Laboratory	DMP	DMTP	DMDTP	DEP	DETP	DEDTP
<b>A</b>	0.6	0.5	0.2	0.3	0.3	0.2
<b>B</b>	5	5	10	5	5	10
<b>C</b>	1.6	1	0.8	1	0.9	0.6
<b>D</b>	2.5	2.5	2.5	2.5	2.5	NA <sup>a</sup>

<sup>a</sup> Lab D does not routinely analyze DEDTP

The study design addressed the principal statistical objective of the project, which was to make statistical comparisons of average analytical results across analysis methods for each of the six compounds. To the extent possible, the design took into account other factors that could have contributed to differences among the analytical results, such as having different participating laboratories and having samples with different concentration levels of the compounds, so that differences among analytical methods could be detected with greater sensitivity.

The study design required each laboratory to analyze 35 samples, with each of the 25 possible sample types represented by either one or two samples. These 35 samples consisted of the following:

- 2 samples where neither Mix A nor Mix B was present
- 4 samples where Mix A was not spiked, but Mix B was spiked at one of three spiking levels (2 samples spiked with Mix B at the Low #2 concentration level; 1 sample spiked with Mix B at each of the medium and high levels)
- 4 samples where Mix B was not spiked, but Mix A was spiked at one of four spiking levels (1 sample spiked with Mix A at each of the Low #1, Low #2, medium, and high spiking levels)
- 7 samples where one mix was spiked at the Low #1 level (Low #2 level for Mix B) and the other mix was spiked at either the Low #2, Medium, or High level (1 sample for each of these 7 spiking combinations)
- 18 samples where each mix was spiked at either the Low #2, Medium, or High spiking levels (2 samples for each of these 9 spiking combinations)

This information on the numbers and types of samples that each laboratory analyzed is summarized within the matrix in Table 4. Each laboratory was directed to test the 35 samples in the order given in Table 5. The testing order was determined by dividing the 35 samples into two groups of 17 plus one extra unspiked sample. Within each group of 17, one sample was tested for each of the 9 sample types where both mixes were represented at either the Low #2, Medium, or High spiked levels. In addition, the samples represented by asterisks within the matrix in Table 4 were tested within the first group of 17 samples. The concentrations of alkyl phosphates corresponding to the Low #1, Low #2, Medium, and High spiking levels are listed in Table 2. Appendix Table B-5 lists the number of samples of each concentration level analyzed as part of this study.

Occasionally, the labs provided more than one measurement for a given urine sample, representing duplicate sample analysis. The statistical analysis included all reported measurements except the unspike pooled urine samples which were omitted so the performance of the methods near their reported detection limits could be evaluated in the absence of background levels of the target analytes. It took into account when measurements were associated with a common sample and the fact that Lab A analyzed three sets of 35 samples.

**Table 4. Sample testing matrix for each participating laboratory, according to the spiking concentration levels associated with each compound mix**

Spiking Levels for Each Mix		Mix B				
		Unspiked	Low Level #1 <sup>a</sup>	Low Level #2	Medium Level	High Level
Mix A	Unspiked	2 samples*	1 sample	1 sample*	1 sample	1 sample*
	Low #1 Level	1 sample	1 sample*	1 sample	1 sample*	1 sample
	Low #2 Level	1 sample*	1 sample	2 samples	2 samples	2 samples
	Medium Level	1 sample	1 sample*	2 samples	2 samples	2 samples
	High Level	1 sample*	1 sample	2 samples	2 samples	2 samples

<sup>a</sup> Due to a sample preparation error, Mix B Low Level # 1 was prepared at the same concentration of Mix B Low Level #2.

\* Included among the first 17 samples tested at each laboratory.



**Table 5. The randomized order of sample testing specified for each laboratory**

Test Number	Sample Type		Test Number	Sample Type	
	Mix A	Mix B <sup>a</sup>		Mix A	Mix B
1 <sup>b</sup>	High	Unspiked	19	Low #1	Low #2
2	Low #2	Low #2	20	Low #2	Medium
3 <sup>b</sup>	Unspiked	Unspiked	21	Low #1	Unspiked
4 <sup>b</sup>	High	High	22	High	Low #2
5	Low #2	Unspiked	23	Low #1	High
6 <sup>b</sup>	Unspiked	High	24 <sup>b</sup>	Unspiked	Medium
7	Medium	Low #1	25 <sup>b</sup>	High	Medium
8	Low #1	Medium	26	Medium	Low #2
9 <sup>b</sup>	Medium	Medium	27 <sup>b</sup>	High	High
10	Low #2	High	28	High	Low #1
11	Unspiked	Low #2	29 <sup>b</sup>	Medium	Unspiked
12	Low #2	Medium	30 <sup>b</sup>	Medium	High
13	High	Low #2	31	Low #2	High
14 <sup>b</sup>	Medium	High	32 <sup>b</sup>	Medium	Medium
15	Low #1	Low #1	33	Unspiked	Low #1
16	Medium	Low #2	34	Low #2	Low #2
17 <sup>b</sup>	High	Medium	35	Unspiked	Unspiked
18	Low #2	Low #1			

<sup>a</sup> Due to a sample preparation error, Mix B Low Level # 1 was prepared at the same concentration of Mix B Low Level #2.

<sup>b</sup> Samples prepared in pooled urine, other samples prepared in synthetic urine

## **Chapter 4**

### **Results and Discussion**

Each participating laboratory submitted the results of their analyses for statistical evaluation. The primary statistical objectives of this study were: 1) to identify those compounds where statistically significant differences existed (at the 0.05 level) in (log-transformed) reported measurements between the concentration levels within each laboratory in order to determine the approximate detection threshold of each laboratory and compare that against their reported detection limits; and 2) to determine when statistically significant differences existed in the (log-transformed) reported measurements between the analytical methods, in order to compare the overall performance of the participating laboratories. To satisfy these objectives, an analysis of variance (ANOVA) model was derived and fitted to the reported measurements. The ANOVA model was fitted separately for each of the six compounds. Appendix B gives a detailed description of the ANOVA model used to analyze the data. The analysis utilized Version 8, Release 8.2, of the SAS<sup>®</sup> System. Measurements falling below a laboratory's detection limit were replaced by one-half of the detection limit prior to the statistical analysis.

#### **Statistical Differences Between Concentration Levels**

Before directly comparing the performance of each laboratory, the approximate detection limit for each compound was determined from the results provided by each laboratory. For example, if measurements for samples spiked at the Low #1 and Low #2 concentrations were found to be statistically equivalent to measurements for unspiked samples, but measurements for samples spiked at the Medium level were found to differ significantly from the unspiked and low-spiked samples, then the first concentration level that should be used to compare laboratory performance should be the Medium level. In the above example, the concentrations reported for the unspiked, Low #1, and Low #2 samples are statistically equivalent to non-detectable results, and therefore, any observed differences between laboratories at these spiking levels are statistically inconsequential.

To address the first statistical objective, statistical tests were performed within the ANOVA to determine, within each laboratory, if significant differences existed in the reported concentrations between the different levels of fortification. When the ANOVA determined that the effect of the spiking concentration was significant (i.e., there were statistically significant differences in the reported results between samples spiked at different concentration levels), then multiple comparison procedures were performed within the ANOVA to determine which pairs of concentration levels differed significantly. Each pairwise comparison of concentration levels of the mix was performed using the Bonferroni-adjustment method, to ensure that the overall error rate associated with all pairwise comparisons was no greater than 0.05. Table 6 displays the results of these statistical tests. Each cell within the table, corresponding to a given compound and laboratory, lists those pairs of concentration levels that are statistically different from each other at an overall 0.05 significance level. For example, if the High level was determined to be significantly different from each of the unspiked, Low #1, Low #2, and Medium levels, the table would show: H vs U, L1, L2, M. This model only reports incidences of significant differences among pairs of spiking levels. Appendix

Tables B-4a through B-4f provide the geometric means of the reported concentrations at each spiking level for each laboratory, as well as other statistical summary parameters that characterize the distribution of the reported data at a given spiking level.

**Table 6. Summary of concentration level effects for each lab and for each alkyl phosphate target compound<sup>a,b</sup>**

Lab	Significant Concentration Level Effect					
	DMP	DMDTP	DEP	DEDTP	DMTP	DETP
Lab A	H vs. U, L1, L2	H vs. U, L1, L2, M	H vs. U, L1, L2, M	H vs. U, L1, L2, M	H vs. U, L2, M	H vs. U, L2, M
	M vs. L1, L2, U	M vs. U, L1, L2	M vs. U, L1, L2	M vs. U, L1, L2	M vs. U, L2	M vs. U, L2
		L2 vs U	L2 vs U	L2 vs. U, L1		L2 vs. U
		L1 vs U	L1 vs U	L1 vs. U		
Lab B	No significant differences.	H vs. U, L1, L2, M	H vs. U, L1, L2, M	H vs. U, L1, L2, M	H vs. U, L2, M	H vs. U, L2, M
		M vs. U, L1, L2	M vs. U, L1, L2	M vs. U, L1, L2	M vs. U, L2	M vs. U, L2
Lab C	H vs. L1, L2, M	H vs. U, L1, L2, M	H vs. U, L1, L2, M	H vs. U, L1, L2, M	H vs. U, L2, M	H vs. U, L2, M
		M vs. U, L1, L2	M vs. U, L1, L2	M vs. U, L1, L2	M vs. U, L2	M vs. U, L2
				L2 vs. U	L2 vs. U	L2 vs. U
Lab D	No significant differences.	H vs. U, L1, L2,	H vs. U, L1, L2	NR <sup>c</sup>	H vs. U, L2, M	H vs. U, L2, M
		M vs. U, L1, L2	M vs. U, L1, L2		M vs. U, L2	M vs. U, L2
					L2 vs. U	L2 vs. U

<sup>a</sup> F tests were used to test for significant concentration level effects for each lab, where the Benjamini and Hochberg multiple comparison adjustment method was used to control the overall error rate across all of these tests to be no higher than 0.05. When significant differences among concentration levels were present for a given lab, pairwise comparisons were made between each pair of concentration levels for the given lab, with each pairwise comparison performed using Bonferroni-adjustment method to ensure that the overall error rate across the pairwise comparisons was no greater than 0.05. Pairs of concentration levels differing significantly at the Bonferroni-adjusted 0.05 level are identified in parentheses.

<sup>b</sup> Mix A compounds and Mix B compounds were spiked at five and four concentration levels, respectively.

<sup>c</sup> No results because this laboratory does not routinely measure DEDTP.

Few significant differences between spiking levels were observed in the reported results for DMP among the laboratories indicating analytical difficulties with this compound (Table 6). The Lab A results indicated that only the Medium and High concentration level were significantly different from (and higher than) the unspiked sample, but that they were not statistically different from one another. The results from Lab C indicate that while results for the High level were significantly different from (and higher than) the Medium, Low #1, and Low #2 levels, results for the unspiked level were not significantly different from any of the spiking levels, including the High level. Lab B and D results indicated no significant differences between the five concentration levels. In addition to these findings, the data in Appendix Table B-4a shows that across all laboratories, the highest individual sample result reported for the High level spike of DMP was 17.5 µg/L, when the known spiked concentration was 200 µg/L. Similarly, for the Medium level samples, the highest individual sample result reported was 15.1 µg/L when the known spiked concentration was 50 µg/L. It is unclear why the results for DMP were so poor across all four laboratories. The addition of DMP to the spiked urine samples was confirmed by review of the solution preparation records. Apparently, a matrix interference in the urine samples may keep the DMP from being extracted and analyzed by either a physical occlusion or a degradation that takes place in the urine matrix. Also, verification of the purity of the DMP standard used to make the urine solutions needs to be done to investigate the possibility that an impure standard was the reason for the poor results for DMP.

Beyond DMP, the interpretation of results in Table 6 for the rest of the compounds is relatively straightforward. For all four laboratories, results for samples spiked at the High and Medium levels were significantly different from (and greater than) the results at the Low and unspiked levels. Below the Medium concentration level the results from each laboratory differed because of range of detection limits for each target analyte.

The reported detection limits for Lab A were all below the Low #1 concentration level. For DMDTP and DEP, Lab A determined the Low #1 and Low #2 concentration levels to be significantly different from the unspiked samples, but was unable to detect a significant difference between the Low #1 and Low #2 concentration levels. For DEDTP, DETP, and DEP, Lab A determined each possible concentration level to be significantly different from the unspiked samples and each other. For DMTP, Lab A was unable to detect a significant difference between the Medium and Low #2 concentration levels. The reported detection limits for Lab B were between the Low #2 and Medium concentration levels. As expected, Lab B was unable to detect a significant difference between any of the Low #1 or #2 concentration levels and the unspiked samples. The detection limits for Labs C and D were at or near the Low #2 concentration level and both performed similarly. For DMDTP and DEP, neither of these labs were able to detect a significant difference between the Low #2, Low #1, or unspiked concentration levels. Also, they were unable to detect a significant difference between the Medium and High concentration levels for those two target analytes. For DEDTP (Lab C only), DMTP, and DETP they both determined each concentration level to be significantly different from the unspiked samples and each other. With the exception of DMP, Table 7 gives the lowest spiked level for each target analyte that was significantly different from the unspiked level.

In summary, Lab A was able to detect concentrations near its reported detection limits to be significantly different from the unspiked samples for all the target analytes except for DMTP.

However, Lab A was able to detect a significant difference between the Low #1 and Low #2 spiking levels for only DEDTP (the Mix B compounds were not spiked at the Low #1 level). Lab B was able to detect concentrations well above its reported detection limits to be significantly different from the unspiked samples for all the target analytes. Labs C and D were able to detect concentrations near its reported detection limits to be significantly different from the unspiked samples for all the target analytes except for DMDTP and DEP (Lab D does not measure DEDTP).

**Table 7. Lowest spiking level of alkyl phosphate target analytes that were significantly larger than the unspiked level (µg/L)<sup>a</sup>**

	DMDTP	DEP	DEDTP	DMTP	DETP
<b>Lab A</b>	Low #1 <sup>b</sup>	Low #1 <sup>b</sup>	Low #1	Medium	Low #2
<b>Lab B</b>	Medium	Medium	Medium	Medium	Medium
<b>Lab C</b>	Medium	Medium	Low #2	Low #2	Low #2
<b>Lab D</b>	Medium <sup>c</sup>	Medium <sup>c</sup>	NR <sup>d</sup>	Low #2	Low #2

<sup>a</sup> All spike levels of DMP were statistically indistinguishable from the unspiked level

<sup>b</sup> Result not statistically different from the Low #2 concentration level

<sup>c</sup> Result not statistically different from the High concentration level

<sup>d</sup> No results because this laboratory does not routinely measure DEDTP.

### Accuracy of Reported Concentrations

An estimate of the accuracy of the results reported by each laboratory relative to the spiking level of the sample (Table 5) was calculated as follows:

$$\% \text{ Accuracy} = 100 \left( \frac{\bar{x}}{T} \right)$$

where x is the mean measured value (across all reported measurements for a given spiking level by a given laboratory), and T is the known fortified concentration. Table 8 gives the average recoveries for concentration levels detectable significantly above the unspiked samples for each lab. Table 9 summarizes the average recoveries listed in Table 8 by providing the range of these average recoveries along with their mean and standard deviation. On average, Labs A and C were within 12% of the known concentrations spiked into the urine samples. They also had relatively small uncertainties around their average recoveries, but the range of their recoveries were from 60% to 180% for Lab A and from 65% to 165% for Lab C. The accuracy achieved by Lab D was approximately 100% on average, but its standard deviation is somewhat larger than Labs A and C. The larger uncertainty is driven by the broad range of recoveries, from 31% to 236%. Lab B grossly over-recovered the target analytes in most instances, which resulted in a

**Table 8. Recoveries of alkyl phosphate target analytes in spiked urine samples**

Analyte	Lab	% Recovery (number of samples) at four different concentration levels			
		Low #1	Low #2	Medium	High
DMP	Lab A	- <sup>a</sup>	-	-	-
	Lab B	-	-	-	-
	Lab C	-	-	-	-
	Lab D	ND <sup>b</sup>	ND	ND	-
DMDTP	Lab A	167.4 (9)	112.4 (8)	74.4 (16)	74.9 (24)
	Lab B	ND	-	1389 (7)	1292 (8)
	Lab C	-	-	85.3 (8)	81.5 (8)
	Lab D	ND	ND	69.0 (8)	52.7 (8)
DEP	Lab A	177.9 (9)	121.2 (16)	64.8 (24)	60.4 (24)
	Lab B	-	-	95.7 (8)	110.8 (8)
	Lab C	-	-	66.3 (8)	64.9 (8)
	Lab D	ND	ND	45.3 (8)	31.4 (8)
DEDTP	Lab A	120.6 (8)	84.9 (23)	81.3 (24)	85.3 (24)
	Lab B	-	-	414.1 (8)	430.1 (8)
	Lab C	-	164.8 (8)	72.1 (8)	79.3 (8)
	Lab D	NR <sup>c</sup>	NR	NR	NR
DMTP	Lab A	NS <sup>d</sup>	-	87.3 (24)	86.4 (24)
	Lab B		-	233.1 (8)	128.0 (8)
	Lab C		84.6 (13)	85.7 (8)	86.7 (8)
	Lab D		235.6 (8)	135.0 (8)	85.4 (8)
DETP	Lab A	NS <sup>d</sup>	180.0 (39)	88.1 (24)	79.2 (24)
	Lab B		-	190.6 (7)	110.7 (8)
	Lab C		103.0 (13)	82.3 (8)	82.0 (8)
	Lab D		156.6 (11)	116.1 (8)	74.1 (8)

<sup>a</sup> The dash indicates that results for this spiking level were not significantly different from the unspiked level for this lab, as indicated in Table 6.

<sup>b</sup> “ND” indicates that at this concentration level, the laboratory reported all the results to be below their detection limit.

<sup>c</sup> No results reported because the laboratory does not routinely measure DEDTP.

<sup>d</sup> “NS” There was no Low #1 level for Mix B.

very large average recovery. Furthermore, recoveries for Lab B ranged from 96% through 1,389%. Overall, the accuracy of Labs A, C, and D were reasonable for this type of analysis, but the rather poor precision across all the laboratories indicates the difficulty in extracting these target analytes from urine in a consistent fashion. The two labs that use isotopically labeled internal standards, Labs A and C, produced more precise results than the other laboratories.

**Table 9. Range of and average recoveries across all target analytes and all detectable concentration levels**

	Range (%)	Average Recovery $\pm$ Standard Deviation (%)
Lab A	60 - 180	103 $\pm$ 39
Lab B	96 - 1,389	439 $\pm$ 490
Lab C	65 - 165	88 $\pm$ 25
Lab D	31 - 236	100 $\pm$ 62

To summarize the performance of each laboratory near their reported detection limits, as discussed in the previous section, Table 10 lists the lowest spiking concentrations that were statistically different from the unspiked samples, the average percent accuracy at that concentration, and the detection limit for the target analytes at each laboratory. It shows the difference between the lowest detectable concentration and the reported detection limit and also how accurate the measurements were at that concentration level.

**Table 10. Lowest spiked concentration (LSC) of alkyl phosphate target analytes that were significantly larger than the unspiked level, the average percent recovery at that spiking level in parentheses, and the reported detection limit (DL) for each participating laboratory (all concentrations in  $\mu\text{g/L}$ )<sup>a</sup>**

	DMDTP		DEP		DEDTP		DMTP		DETP	
	LSC	DL	LSC	DL	LSC	DL	LSC	DL	LSC	DL
Lab A	1.04 <sup>b</sup> (167)	0.2	1.40 <sup>b</sup> (121)	0.3	1.35 (121)	0.2	23.9 (87)	0.5	4.44 (180)	0.3
Lab B	52.0 (1,389)	10	70.0 (96)	5	67.5 (414)	10	23.9 (233)	5	27.8 (191)	5
Lab C	52.0 (85)	0.8	70.0 (66)	1	2.70 (165)	0.6	3.82 (85)	1	4.44 (103)	0.9
Lab D	52.0 <sup>c</sup> (69)	2.5	70.0 <sup>c</sup> (45)	2.5	NR <sup>d</sup>	NA <sup>a</sup>	3.82 (236)	2.5	4.44 (157)	2.5

<sup>a</sup> All spike levels of DMP were statistically indistinguishable from the unspiked level

<sup>b</sup> Result not statistically different from the Low #2 concentration level

<sup>c</sup> Result not statistically different from the High concentration level

<sup>d</sup> No results because this laboratory does not routinely measure DEDTP.

### **Statistical Differences Between the Laboratories**

The second statistical objective was to investigate the presence of significant differences in the reported results among laboratories, taking into account the different spiking levels of the samples. This objective was addressed by performing additional statistical tests within the ANOVA discussed earlier. When statistical tests determined that the laboratory effect was significant at a given spiking level (i.e., there were statistically significant differences in the reported results among laboratories), then those laboratories whose results at that spiking level were significantly different from the unspiked level (Table 6) were identified. Among these laboratories, those pairs of

laboratories that differed significantly at that spiking level were identified. Each pairwise comparison of laboratories was performed using the Bonferroni - adjustment method, to ensure that the overall error rate associated with all pairwise comparisons (at a given spiking level) was no greater than 0.05 (Table 11). Because the Low #1 spiking level was found to differ significantly from the unspiked level only for Lab A, that concentration level was omitted from the table. Similarly, no results are included in Table 11 for DMP, as no spiking levels differed significantly from the unspiked level for any laboratory. The laboratories included in the pairwise comparisons are noted within each cell.

For DMDTP at the Medium and High spiking levels, Lab B, with average recoveries of 1,389% and 1,292%, was significantly different from the other three laboratories, whose average recoveries ranged from 53% to 85%. Labs A, C, and D did not differ significantly from each other at these two spiking levels.

For DEP, there was no significant difference between laboratories at the Medium concentration level according to the ANOVA, so no pairwise comparisons of labs were performed. The average recoveries for all the laboratories ranged from 45% to 96%. At the High concentration level, where all four laboratories were also compared, Lab B (111% average recovery) differed significantly only from Lab D (31% average recovery), while Labs A (60% average recovery), C (65% average recovery), and D did not differ significantly from one another. Also, Lab B did not differ significantly from Labs A and C.

For DEDTP at the Low #2 concentration level, only Labs A and C were compared, and there was no significant difference found between them. Their average recoveries were 85 and 165%, respectively. At the Medium and High concentration levels, when pairs of all three laboratories analyzing DEDTP were compared, Lab B, which had average recoveries of 414 and 430%, was significantly different from Labs A and C, whose average recoveries ranged from 72 to 85%. Labs A and C were not significantly different from one another.

For DMTP, only Labs C and D were compared at the Low #2 concentration level and there was no significant difference found. Their average recoveries were 85%, and 236%, respectively. All four laboratories were compared at the two higher concentration levels. At the Medium level, Lab A (87% average recovery) was significantly different from Lab B (233% average recovery) but was not significantly different from Labs C (86% average recovery) and D (135% average recovery). Labs B, C, and D were not significantly different from one another. At the High level, there was no significant difference among the four laboratories; their average recoveries ranged from 85% to 128%.



**Table 11. Summary of lab effects at low #2, medium, and high concentration levels <sup>a</sup>**

Concentration Level	Significant Lab Effects					
	DMP	DMDTP	DEP	DEDTP <sup>d</sup>	DMTP	DETP
Low #2	NA <sup>b</sup>	- <sup>c</sup>	-	No significant pairs of differences (comparing Labs A and C only)	No significant pairs of differences (comparing Labs C and D only)	Lab A vs Lab C (comparing Labs A, C, and D only)
Medium	NA	Lab B vs Lab A, Lab C, Lab D (comparing all labs)	No significant differences in labs	Lab B vs Lab A, Lab C (comparing all labs)	Lab A vs Lab B (comparing all labs)	No significant differences in labs
High	NA	Lab B vs Lab A, Lab C, Lab D (comparing all labs)	Lab B vs Lab D (comparing all labs)	Lab B vs Lab A, Lab C (comparing all labs)	No significant differences in labs	No significant differences in labs

<sup>a</sup> F tests were used to test for significant lab effects at each concentration level of the compound, where the Benjamini and Hochberg multiple comparison adjustment method was used to control the overall error rate across all of these tests to be no higher than 0.05. When significant differences among labs were present at a given concentration level, pairwise comparisons were made between each pair of labs at the given concentration level, with each pairwise comparison performed using a Bonferroni-adjustment method to ensure that the overall error rate across the pairwise comparisons was no greater than 0.05. Pairs of labs differing significantly at the Bonferroni-adjusted 0.05 level are identified in parentheses.

<sup>b</sup> No labs had spiking levels that differed significantly from the unspiked level for DMP (see Table 6).

<sup>c</sup> The dash indicates that the specified spiking level differed significantly from the unspiked level for either no lab or only one lab, and so no pairwise comparisons are reported among labs.

<sup>d</sup> Lab D did not analyze for DEDTP.

For DETP, at the Low #2 concentration level, Lab A (180% average recovery) was significantly different from Lab C (103% average recovery) but was not significantly different from Lab D (157% average recovery). Labs C and D were not significantly different from one another. At the Medium and High concentration levels, there were no significant difference among the four laboratories; their average recoveries ranged from 82% to 191% at the Medium level, and from 74% to 111% at the High level.

Overall, Lab A reported concentrations that were significantly different from (and higher than) the unspiked samples for DMDTP, DEP, and DEDTP at the Low #1 and #2 concentration levels, DETP at the Low #2 concentration level, and for all of the target analytes except DMP at the Medium and High concentration levels. The range of average recoveries for Lab A was 60% to 180%. Lab B reported concentrations that were significantly different from (and higher than) the unspiked samples for all the target analytes except DMP at the Medium and High concentration levels. However, with the exception of DEP and the High level concentrations of DMTP and DETP, Lab B's recoveries were generally much greater than 100%. Lab C reported concentrations that were significantly different from the unspiked samples for DEDTP, DMTP, and DETP at the Low #2 concentration level and for all the target analytes except DMP at the Medium and High

concentration levels. The range of average recoveries for Lab C was 65% to 165%. Lab D reported concentrations that were significantly different from the unspiked samples for DMTP and DETP at the Low #2 concentration level, and for all of the target analytes except DMP at the Medium and High concentration levels. However, for DMDTP and DEP, the High level concentrations were not statistically different from the Medium level. The range of average recoveries for Lab D was 31% to 236%.

## References

- Bradway, D.E., Shafik, T.M., and Lores, E.M. (1977). "Comparison of cholinesterase activity, residue levels, and urinary metabolite excretion of rats exposed to organophosphorus pesticides." *J. Agric. Food Chem.*, 25:1353-1358.
- Fortmann, R.C., Sheldon, L.S., Smith, D., Perritt, K., and Camann, D.E. (1991). House Dust/Infant Pesticides Exposure Study (HIPES). Final report to EPA, Contract No. 68-02-4544, Research Triangle Institute and Southwest Research Institute.
- Franklin, C.A., Fenske, R.A., Greenhalgh, R., Mathieu, L., Denley, H.V., Leffingwell, J.T., and Spear, R.C. (1981). "Correlation of urinary pesticide metabolite excretion with estimated dermal contact in the course of occupational exposure to Guthion." *J. Toxicol. Environ. Health* 7:715-731.
- Lewis, R.G., Bond, A.E., Johnson, D.E., and Shu, J. P. (1988). "Measurement of atmospheric concentrations of common household pesticides: A pilot study." *Environ. Monitoring Assessment* 10:59-73.
- Lin, D.C.K., Melton, R.G., Kopfler, F.C., and Lucas, S.V. (1981). "Glass capillary gas chromatographic/mass spectrometric analysis of organic concentrates from drinking and advanced waste treatment waters." In: *Advances in the Identification and Analysis of Organic Pollutants in Water*, Vol. 2, Chapter 46 (L.H. Keith, ed.). Ann Arbor Science Publ., Inc., Ann Arbor, MI. pp. 861-906.
- Morgan, D.P., Hetzler, H.L., Slach, E.F., and Lin, L.I. (1977). "Urinary excretion of paranitrophenol and alkyl phosphates following ingestion of methyl or ethyl parathion by human subjects. *Arch. Environ. Contam. Toxicol.* 6:159-173.
- Murphy, R.S., Kutz, F.W., and Strassman, S.C. (1983). "Selected pesticide residues or metabolites in blood and urine specimens from a general population survey." *Environ. Health Perspect.* 48:81-86.
- Schattenberg, H.J., III and Hsu, J.-P. (1992). "Pesticide residue survey of produce from 1989 to 1991." *J. AOAC Intl.* 75:925-933.

## **Appendix A**

### **Descriptions of the Participating Laboratories' Methods for Measuring Alkyl Phosphates in Urine**

#### Lab A

1. 4.0 mL aliquot of urine is spiked with 25 µg/L of deuterated DMP, DMDTP, DEP, DMTP, and DETP and <sup>13</sup>C- labeled DEDTP, as internal standards.
2. 4 mL acetonitrile added to the sample and mixed.
3. Samples evaporated at 50°C using a Turbovap apparatus until approximately 4 mL of solution remains.
4. An additional 4 mL of acetonitrile is added and the evaporation is repeated using the same conditions until 2 mL of solution remained.
5. This step is repeated once more until the urine contents in the tube were totally concentrated.
6. The concentrated residue is reconstituted with 1 mL of acetonitrile and 50 µL of derivatizing agent, 1-chloro-3-iodopropane.
7. The sample is maintained at room temperature for 1 h and then is transferred to a clean test tube. A few grains of potassium carbonate are added to the sample, and the sample is placed in a heater block for 2 h at 80°C.
8. The sample is evaporated in a Turbovap apparatus using the same conditions as described above until the final volume was 100 µL.
9. The sample is transferred to an autosampler vial, sealed, and stored at -20°C until analysis.
10. 1 µL of each sample is injected into a triple quadrupole GC/MS outfitted with a 30-m DB-5MS capillary column (0.25 mm i.d., 0.25 µm film thickness). One quantitation ion is the molecular ion produced by chemical ionization in the positive ion mode and the other quantitation ion is the daughter ion produced by the insertion of a collision induced dissociation gas.

#### Lab B

1. 1 mL of urine is spiked with fenthion as the internal standard
2. Samples are lyophilized and then derivatized with a benzyltolyltriazine reagent
3. A saturated salt solution is added to the reaction vessel and the benzyl derivatives are extracted with cyclohexane.
4. Solution is then analyzed by GC with a flame photometric detector outfitted with a 30-m DB-210 capillary column (0.53 mm i.d., 1.0 µm film thickness).

### Lab C

1. 0.5 mL urine sample spiked with deuterated DMP, DEP, DMTP, and DETP as internal standards.
2. Acetonitrile is added to the sample and the mixture is centrifuged.
3. The supernatant liquid is evaporated to dryness and redissolved in pure acetonitrile.
4. Sample is then derivatized with pentafluorobenzyl bromide along with the addition of potassium carbonate as a catalyst at 70°C for 2 h.
5. The derivatized alkyl phosphates are extracted twice with a mixture of dichloromethane in hexane (8% v/v), filtered on sodium sulphate and evaporated to 0.2 mL.
6. The alkyl phosphates are quantified on GC/MS with electron impact ionization. The GC/MS is outfitted with a 30-m HP-50+ capillary column (0.25 mm i.d., 0.25 µm film thickness).

### Lab D

1. 5 mL of urine is pipetted into a centrifuge tube, 35 mL of acetonitrile is added and the mixture is centrifuged at 2500 rpm for 10 min.
2. The supernatant liquid is decanted into a TurboVap flask and evaporated to 1 mL.
3. Quantitatively transfer the distillate to a 15 mL screw cap test tube and add 1.5 mL of methanol to the TurboVap flask.
4. Add 2 mL acetonitrile to the test tube and a bilayer will form.
5. Quantitatively transfer methanol from TurboVap flask to test tube which will cause the formation of a clear yellow solution and some precipitate. Add 8 mL of acetone. Vortex.
6. Centrifuge tubes for 10 minutes at 2500 rpm.
7. Decant supernatant to a new test tube. Evaporate to dryness under a gentle stream of nitrogen. Sample residue will be approximately 0.25 mL of yellow oil.
8. To the oil residue, add 1 mL dehydrated acetone.
9. Add 20 µL pentafluorobenzyl bromide derivatizing reagent.
10. Cap and rotate at room temperature for 30 min. Evaporate to near dryness under gentle stream of nitrogen.
11. Add approximately 20 mg of potassium carbonate to the dry residue.
12. Extract with 10 mL hexane using vortex.
13. Decant hexane extracts into a TurboVap flask and evaporate to 0.5 mL (thio-containing phosphates).
14. Add an additional 20 mg of potassium carbonate to the dry residue. Add 1.0 mL dehydrated acetonitrile/dimethylformamide (4:1). Pipette 20 µL pentafluorobenzyl bromide derivatizing reagent into the sample. Vortex.
15. Cap and derivatize at 90°C for 30 minutes.
16. Cool samples, add 2 mL water to dissolve remaining potassium carbonate.
17. Extract residue with 3 x 5 mL hexane.
18. Combine extract and evaporate to 0.5 mL for analysis (non-thio-containing alkyl phosphates)
19. Both the thio and non-thio-containing extracts are then analyzed by GC with a pulsed flame photometric detector outfitted with a 30-m SPB-20 capillary column (0.32 mm i.d., 1.0 µm film thickness).

## Appendix B Statistical Methods and Results

### Statistical Analysis Methods

For each of the six alkyl phosphate compounds, the statistical objectives of this study were 1) to identify those laboratories (i.e., analytical methods) where statistically significant differences existed (at the 0.05 level) in reported measurements between spiking levels, and particularly, with the unspiked level, and 2) to identify those spiking levels for which statistically significant differences existed (at the 0.05 level) in reported measurements between the laboratories (i.e., analytical methods). To satisfy these objectives, an analysis of variance (ANOVA) model was derived and fitted to the reported measurements. The model was fitted separately for each compound, corresponding to a total of six model fits. The data analysis utilized Version 8, Release 8.2, of the SAS<sup>®</sup> System.

Note that each laboratory utilized a different analytical method. Thus, this statistical analysis could not distinguish whether observed differences in results between two analytical methods are due to differences associated with the methods or to differences associated with the laboratories performing these methods.

Descriptive statistics of the reported measurements were calculated within tables and figures (plots). These statistics include sample size, geometric means, standard deviation, and selected percentiles. These summaries and other investigations of the reported measurements concluded that the ANOVA model would be fitted to the log-transformed measurements. Measurements which the laboratory reported as zero or less than the detection limit were replaced with one-half of the detection limit prior to summarizing the measurements and analyzing the measurements using ANOVA. However, results that specified a particular value that fell below the detection limit were retained as reported when the data were summarized and analyzed.

The ANOVA model took the following form:

$$\begin{aligned}
 Y_{ijkmrs} = & \mu + \text{LAB}_i + C_j^1 + C_k^0 + (C^1 * C^0)_{jk} + (\text{LAB} * C^1)_{ij} + (\text{LAB} * C^0)_{ik} + (\text{LAB} * C^1 * C^0)_{ijk} \\
 & + \text{SET}_{m(I)} + \text{REP}_{r(ijkm)} + \epsilon_{s(ijkmr)}
 \end{aligned} \tag{1}$$

(i=1,...,I; j=1,...,J; k=1,...,K; m=1,...,M<sub>i</sub>; r=1,...,R<sub>jk</sub>; s=1,...,S<sub>ijkmr</sub>)

where

- $Y_{ijkmrs}$  denotes the log-transformed measurement for the s<sup>th</sup> analysis performed on the physical sample uniquely identified by the combination of subscripts (i,j,k,m,r) (these subscripts are more fully defined in the bullets that follow),
- $\mu$  is an overall constant,

- $LAB_i$  is a fixed effect representing the  $i^{th}$  laboratory or equivalently, the  $i^{th}$  analytical method ( $i=3$  [Lab A, Lab B, Lab C] for DEDTP;  $i=4$  [Lab A, Lab B, Lab C, Lab D] for the other compounds),
- $C_j^1$  is a fixed effect representing the  $j^{th}$  spiking level of the mix in which the given compound is included ( $j=5$  for Mix A [DMP, DMDTP, DEP, and DEDTP, where the spiking levels were denoted by Unspiked, Low #1, Low #2, Medium, High];  $j=4$  for Mix B [DMTP and DETP, where the spiking levels were denoted by Unspiked, Low, Medium, and High]),
- $C_k^0$  is a fixed effect representing the  $k^{th}$  spiking level of the mix not containing the given compound ( $k=5$  if this other mix is Mix A;  $k=4$  if this other mix is Mix B),
- $SET_{m(i)}$  is a random effect representing the  $m^{th}$  set of 35 samples provided to the  $i^{th}$  laboratory ( $m(1)=3$  [i.e., for Lab A];  $m(i)=1$  otherwise),
- $REP_{r(ijkm)}$  is a random effect representing the  $r^{th}$  sample containing the  $j^{th}$  spiking level of the mix in which the given compound is included and the  $k^{th}$  spiking level of the other mix, where the sample is within the  $m^{th}$  set of samples analyzed by the  $i^{th}$  laboratory ( $r=1, 2, \text{ or } 3$ , depending on the sample type defined by the combination  $(j,k)$ ),
- Terms containing asterisks represent interactions of the above effects, and
- $\epsilon_{s(ijkmr)}$  is random error not attributable to the model, as represented by variability in results for duplicate analyses of the same physical sample within a laboratory (where  $s$  can range from 1 to 4, depending on the specific combination of  $(i,j,k,m,r)$ ).

Model (1) was fitted using the MIXED procedure in the SAS<sup>®</sup> System.

Within the fitted ANOVA for each compound, two sets of statistical tests were performed to address the two statistical analysis objectives stated above:

- F-tests for significant differences among spiking levels for the mix containing the given compound, one test for each laboratory.
- F-tests for significant differences among laboratories, one test for each spiking level for the mix containing the given compound.

These tests were possible due to having a term representing the interaction of laboratory and spiking level effects in the model. The significance levels for each F-test in a set (and for a particular compound) were adjusted using the Benjamini and Hochberg method (available in the SAS<sup>®</sup> System), and an adjusted significance level below 0.05 resulted in the given test being declared significant. Thus, any test in Set #1 having an adjusted significance level of less than 0.05 indicated that significant differences among spiking levels existed for the given laboratory, and any test in Set #2 having an adjusted significance level of less than 0.05 indicated that significant differences

among laboratories existed at the given spiking level. This adjustment of significance levels was necessary to ensure that the overall rate of erroneously declaring a given test as significant within a given set of tests (for a given compound) was no higher than 0.05.

When significant differences were observed among spiking levels for a given laboratory (i.e., the outcome of a test in Set #1), additional F-tests were performed within the ANOVA to determine those pairs of spiking levels that differed significantly for that laboratory. A Bonferroni multiple comparisons method was used, indicating that each test needed to have a significance level less than  $0.05/T$ , where T was the total number of pairs of spiking levels, in order for the given pair of spiking levels to be declared significantly different. (T=10 for Mix A compounds, and T=6 for Mix B compounds.) This approach ensured that for a given laboratory, the overall error rate among all T pairwise comparisons was no greater than 0.05.

Similarly, when significant differences were observed among laboratories at a given spiking level (i.e., the outcome of a test in Set #2), F-tests were performed within the ANOVA to determine those pairs of laboratories that differed significantly at that spiking level. When the laboratory effect was significant at a given spiking level, then each pair of laboratories was statistically compared within the ANOVA. Each pairwise comparison of laboratories was performed using the Bonferroni adjustment method, indicating that each test needed to have a significance level less than  $0.05/T$ , where T was the total number of pairs of laboratories of interest, in order for the given pair of laboratories to be declared significantly different. While the analyses presented in this appendix considered all possible pairs of laboratories, the analysis presented in the main body of this report considered only those pairs where both laboratories reported measurements at the given spiking level that were significantly different from (and, on average, higher than) the unspiked level, according to the tests described in the previous paragraph.

### Data Analysis Results

For each compound, Table B-1 specifies the number of measurements reported by each laboratory, by spiking level. In most cases, one measurement was reported for each compound for a given physical sample. However, occasionally the laboratories reported duplicate measurements for the same physical sample for certain compounds. These incidents are noted in parentheses within Table B-1.

For each compound, Table B-2 specifies the number of measurements reported by each laboratory that were below the laboratory's detection limit (given in Table 1 of the main report). The total number of measurements and the percentage of measurements below the detection limit are also specified in these tables. These numbers are reported by laboratory and spiking level and include both individual sample results and duplicate results for the same sample.

For each compound, Table B-3 summarizes accuracy percentages (i.e., average measurement divided by the actual spiking spiking, specified in Table 5 of the main report, and expressed as a percentage) that are associated with the reported measurements that were above the detection limit. These percentages are reported by laboratory and spiking level. Note that no duplicate results for



the same sample, and no non-detected results, were used in calculating the average measurement that is given in the numerator of these accuracy percentages. Cells containing a dash symbol (--) indicate that no measurements were reported above detection limits.

Tables B-4a through B-4f contain descriptive statistics of the reported analytical measurements, presented by laboratory and spiking level, for each set of samples received by a laboratory. Separate tables exist for each compound. Note that the measurements summarized in these tables include duplicate measurements taken on the same physical sample. Measurements reported as zero or below the detection limit were replaced with one-half of the detection limit. Due to the possible of contamination in the pooled urine samples, the following samples were excluded: test numbers 3, 6, and 24 for compounds in Mix A; and test numbers 1, 3, and 29 for compounds in Mix B. Because the data were analyzed after taking log transformations, geometric means and geometric standard deviations, equal to the exponential value of the arithmetic mean and standard deviation of the log-transformed data, are presented in these tables. The geometric means presented in these tables for each spiking level and laboratory are presented graphically in Figures B-1 through B-6, with separate figures for each compound.

**Table B-1. Numbers of Samples with Analytical Measurements Reported for Mix A and Mix B Compounds, by Laboratory and Spiking Level**

Laboratory	Unspiked	Low #1	Low #2	Medium	High	All Samples
<b>Mix A Compounds</b>						
Lab A (Set #1)	6	5	8	8	8 (4 dup. results for 1 sample: DEDTP)	35
Lab A (Set #2)	6 (5 for DMDTP)	5	8 (2 dup. results for 1 sample: DMP)	8 (2 dup. results for 1 sample: DMP)	8	35 (34 for DMDTP)
Lab A (Set #3)	6	5	8 (2 dup. results for 1 sample: DMP)	8	8	35
Lab B	6	5	8	8	8	35
Lab C	6	5 (2 dup. results for 3 samples)	8 (2 dup. results for 1 sample)	8 (2 dup. results for 2 samples)	8 (2 dup. results for 2 samples)	35
Lab D <sup>a</sup>	6	5	8	8 (2 dup. results for 2 samples)	8 (2 dup. results for 1 sample)	35
<b>Mix B Compounds</b>						
Lab A (Set #1)	6	13		8	8	35
Lab A (Set #2)	6	13		8	8	35
Lab A (Set #3)	6	13		8	8	35
Lab B	6	13		8	8	35
Lab C	6 (2 dup. results for 2 samples)	13 (2 dup. results for 1 sample)		8 (2 dup. results for 3 samples)	8 (2 dup. results for 2 samples)	35
Lab D	6 (2 dup. results for 2 samples)	13		8 (2 dup. results for 1 sample)	8	35

<sup>a</sup> No measurements were reported for DEDTP.

**Table B-2. Number of Not-Detected Analytical Measurements for Each Compound, Calculated by Laboratory and Spike Level, with Number of Analytical Measurements and the Not-Detected Percentage Given in Parentheses**

Laboratory	# Not-Detected Measurements (Total # Measurements, % of Measurements that are Not-Detected)					
	Unspiked	Low #1	Low #2	Medium	High	Overall
<b>Compound = DMP</b>						
Lab A	4 (18, 22.2)	6 (15, 40.0)	10 (26, 38.5)	4 (25, 16.0)	6 (27, 22.2)	30 (111, 27.0)
Lab B	6 (6, 100)	4 (5, 80.0)	6 (8, 75.0)	7 (8, 87.5)	2 (8, 25.0)	25 (35, 71.4)
Lab C	3 (6, 50.0)	4 (8, 50.0)	2 (9, 22.2)	3 (10, 30.0)	0 (10, 0.0)	12 (43, 27.9)
Lab D	6 (6, 100)	5 (5, 100)	8 (8, 100)	10 (10, 100)	7 (9, 77.8)	36 (38, 94.7)
<b>Compound = DMDTP</b>						
Lab A	11 (18, 61.1)	7 (15, 46.7)	10 (26, 38.5)	1 (25, 4.0)	3 (27, 11.1)	32 (111, 28.8)
Lab B	5 (6, 83.3)	5 (5, 100)	5 (8, 62.5)	1 (8, 12.5)	0 (8, 0.0)	16 (35, 45.7)
Lab C	5 (6, 83.3)	7 (8, 87.5)	0 (9, 0.0)	0 (10, 0.0)	0 (10, 0.0)	12 (43, 27.9)
Lab D	6 (6, 100)	5 (5, 100)	8 (8, 100)	0 (10, 0.0)	0 (9, 0.0)	19 (38, 50.0)
<b>Compound = DEP</b>						
Lab A	7 (18, 38.9)	6 (15, 40.0)	10 (26, 38.5)	1 (25, 4.0)	3 (27, 11.1)	27 (111, 24.3)
Lab B	5 (6, 83.3)	3 (5, 60.0)	3 (8, 37.5)	0 (8, 0.0)	0 (8, 0.0)	11 (35, 31.4)
Lab C	2 (6, 33.3)	0 (8, 0.0)	0 (9, 0.0)	0 (10, 0.0)	0 (10, 0.0)	2 (43, 4.7)
Lab D	6 (6, 100)	5 (5, 100)	7 (8, 87.5)	0 (10, 0.0)	0 (9, 0.0)	18 (38, 47.4)
<b>Compound = DEDTP</b>						
Lab A	12 (18, 66.7)	7 (15, 46.7)	3 (26, 11.5)	1 (25, 4.0)	0 (27, 0.0)	23 (111, 20.7)
Lab B	4 (6, 66.7)	4 (5, 80.0)	3 (8, 37.5)	0 (8, 0.0)	0 (8, 0.0)	11 (35, 31.4)
Lab C	5 (6, 83.3)	0 (8, 0.0)	0 (9, 0.0)	0 (10, 0.0)	0 (10, 0.0)	5 (43, 11.6)
<b>Compound = DMTP</b>						
Lab A	0 (18, 0.0)	4 (42, 9.5)		2 (26, 7.7)	1 (25, 4.0)	7 (111, 6.3)
Lab B	3 (6, 50.0)	6 (13, 46.2)		0 (8, 0.0)	0 (8, 0.0)	9 (35, 25.7)
Lab C	4 (8, 50.0)	0 (14, 0.0)		0 (11, 0.0)	0 (10, 0.0)	4 (43, 9.3)
Lab D	5 (8, 62.5)	5 (13, 38.5)		0 (9, 0.0)	0 (8, 0.0)	10 (38, 26.3)
<b>Compound = DETP</b>						
Lab A	4 (18, 22.2)	3 (42, 7.1)		2 (26, 7.7)	1 (25, 4.0)	10 (111, 9.0)
Lab B	3 (6, 50.0)	6 (13, 46.2)		1 (8, 12.5)	0 (8, 0.0)	10 (35, 28.6)
Lab C	2 (8, 25.0)	0 (14, 0.0)		0 (11, 0.0)	0 (10, 0.0)	2 (43, 4.7)
Lab D	6 (8, 75.0)	2 (13, 15.4)		0 (9, 0.0)	0 (8, 0.0)	8 (38, 21.1)

Note: "Not-detected measurements" are any measurements that fall below a laboratory's reported detection limit for the given compound.

**Table B-3. Accuracy Estimates (%) for Each Compound, Calculated by Laboratory and Spike Level, with Number of Analytical Measurements Falling Above the Detection Limit Given in Parentheses**

Laboratory	Spiking Level (# Measurements > Detection Limit)			
	Low #1	Low #2	Medium	High
<b>Compound = DMP</b>				
Lab A	167.4 (9)	127.6 (14)	11.7 (20)	3.5 (21)
Lab B	1190 (1)	492.5 (2)	17.4 (1)	4.9 (6)
Lab C	423.7 (3)	185.5 (6)	7.8 (6)	3.6 (8)
Lab D	–	–	–	2.6 (2)
<b>Compound = DMDTP</b>				
Lab A	112.4 (8)	74.4 (16)	70.8 (24)	74.9 (24)
Lab B	–	4635 (3)	1389 (7)	1292 (8)
Lab C	76.9 (1)	73.0 (8)	85.3 (8)	81.5 (8)
Lab D	–	–	69.0 (8)	52.7 (8)
<b>Compound = DEP</b>				
Lab A	177.9 (9)	121.2 (16)	64.8 (24)	60.4 (24)
Lab B	535.7 (2)	742.9 (5)	95.7 (8)	110.8 (8)
Lab C	151.1 (5)	109.1 (8)	66.3 (8)	64.9 (8)
Lab D	–	235.7 (1)	45.3 (8)	31.4 (8)
<b>Compound = DEDTP</b>				
Lab A	120.6 (8)	84.9 (23)	81.3 (24)	85.3 (24)
Lab B	10141 (1)	1518 (5)	414.1 (8)	430.1 (8)
Lab C	88.6 (5)	164.8 (8)	72.1 (8)	79.3 (8)
<b>Compound = DMTP</b>				
Lab A	78.0 (38)		87.3 (24)	86.4 (24)
Lab B	650.0 (7)		233.1 (8)	128.0 (8)
Lab C	84.6 (13)		85.7 (8)	86.7 (8)
Lab D	235.6 (8)		135.0 (8)	85.4 (8)
<b>Compound = DETP</b>				
Lab A	180.0 (39)		88.1 (24)	79.2 (24)
Lab B	355.5 (7)		190.6 (7)	110.7 (8)
Lab C	103.0 (13)		82.3 (8)	82.0 (8)
Lab D	156.6 (11)		116.1 (8)	74.1 (8)

Note: Accuracy is estimated by (mean/T)\*100%, where “mean” is the arithmetic mean of the analytical measurements falling above the detection limit, calculated across all samples spiked at the specified level, and T is the actual spiking level.

**Table B-4a. Descriptive Statistics of Reported Analytical Measurements for DMP (µg/L), Calculated by Spiking Level for Each Laboratory and Across All Laboratories**

Lab	Set	Spiking Level	# Measurements	Geom. Mean	Geom. Standard Deviation	Minimum	25 <sup>th</sup> Percentile	Median	75 <sup>th</sup> Percentile	Maximum
Lab A	1	Unspiked	3	0.9	2.6	0.3	0.3	1.4	1.7	1.7
		Low #1	5	0.6	2.3	0.3	0.3	0.5	1	2
		Low #2	8	1.2	2.8	0.3	0.5	1.6	1.8	7.8
		Medium	8	2.5	3.8	0.3	1.2	2.7	8.1	15.8
		High	8	3.4	3.5	0.3	1.5	4.3	10.1	14.7
	2	Unspiked	3	1.3	1.3	1	1	1.3	1.5	1.5
		Low #1	5	1.5	1.3	1.1	1.2	1.5	1.7	2.3
		Low #2	9	1.3	3.3	0.3	0.3	1.7	3.2	5.7
		Medium	9	2.7	3	0.3	1.8	2.6	3.4	13.8
		High	8	3.5	3.5	0.3	2	4.6	9.7	13.5
	3	Unspiked	3	0.3	1	0.3	0.3	0.3	0.3	0.3
		Low #1	5	0.7	2.7	0.3	0.3	0.5	1.6	2.5
		Low #2	9	0.8	2.8	0.3	0.3	0.9	2.1	3
		Medium	8	2.4	4.4	0.3	1	3.1	8.7	15.1
		High	8	3.9	3.8	0.3	1.9	5.3	11.4	17.5
	Overall	Unspiked	9	0.7	2.2	0.3	0.3	1	1.4	1.7
		Low #1	15	0.9	2.2	0.3	0.3	1.1	1.7	2.5
		Low #2	26	1.1	2.9	0.3	0.3	1.6	2.1	7.8
		Medium	25	2.6	3.5	0.3	1.7	2.8	3.8	15.8
		High	24	3.6	3.4	0.3	1.5	4.5	9.7	17.5
Lab B	1	Unspiked	3	3.1	1.4	2.5	2.5	2.5	4.6	4.6
		Low #1	5	3.4	2	2.5	2.5	2.5	2.5	11.9
		Low #2	8	3.7	1.8	2.5	2.5	2.8	4.9	13.3
		Medium	8	3.2	1.6	2.5	2.5	2.5	3.7	8.7
		High	8	7.4	1.7	2.5	5.4	9.3	11.2	11.8
Lab C	1	Unspiked	3	4.3	18.1	0.8	0.8	0.8	120.6	120.6
		Low #1	8	1.8	2.2	0.8	0.9	1.9	3.6	5.7
		Low #2	9	2.6	2.1	0.8	2.1	3.3	4.5	5.9
		Medium	10	2.1	2.2	0.8	1	2.2	3.6	7.3
		High	10	6.8	1.4	3.2	6	7	8.3	10.3
Lab D	1	Unspiked	3	1.3	1	1.3	1.3	1.3	1.3	1.3
		Low #1	5	1.3	1	1.3	1.3	1.3	1.3	1.3

		Low #2	8	1.3	1	1.3	1.3	1.3	1.3	1.3
		Medium	10	1.3	1	1.3	1.3	1.3	1.3	1.3
		High	9	1.7	1.9	1.3	1.3	1.3	1.3	5.9
All Labs	Overall	Unspiked	18	1.3	4	0.3	0.8	1.3	1.7	120.6
		Low #1	33	1.4	2.4	0.3	1	1.3	2.5	11.9
		Low #2	51	1.6	2.6	0.3	0.9	1.7	3.1	13.3
		Medium	53	2.2	2.7	0.3	1.3	2.5	3.4	15.8
		High	51	4	2.8	0.3	1.5	5.3	8.3	17.5

Note: In calculating these statistics, results reported as “below detection limits” were replaced by one-half of the laboratory’s detection limit.

**Table B-4b. Descriptive Statistics of Reported Analytical Measurements for DMDTP ( $\mu\text{g/L}$ ), Calculated by Spiking Level for Each Laboratory and Across All Laboratories**

Lab	Set	Spiking Level	# Measurements	Geom. Mean	Geom. Standard Deviation	Minimum	25 <sup>th</sup> Percentile	Median	75 <sup>th</sup> Percentile	Maximum
Lab A	1	Unspiked	3	0.1	1.0	0.1	0.1	0.1	0.1	0.1
		Low #1	5	0.3	3.7	0.1	0.1	0.1	1.0	1.2
		Low #2	8	0.4	4.6	0.1	0.1	0.5	1.8	2.5
		Medium	8	36.6	1.1	35.1	35.4	35.8	37.4	41.1
		High	8	158.4	1.0	154.4	156.5	157.9	161.0	162.4
	2	Unspiked	3	0.2	3.4	0.1	0.1	0.1	0.8	0.8
		Low #1	5	0.4	3.4	0.1	0.1	0.8	0.8	1.2
		Low #2	8	1.0	2.5	0.1	1.1	1.4	1.5	1.5
		Medium	8	38.0	1.1	33.5	36.7	38.7	39.9	40.2
		High	8	156.8	1.1	139.2	154.6	158.1	161.8	167.8
	3	Unspiked	3	0.1	1.0	0.1	0.1	0.1	0.1	0.1
		Low #1	5	0.5	4.3	0.1	0.1	0.8	1.4	2.2
		Low #2	8	0.6	4.3	0.1	0.1	1.5	1.7	2.1
		Medium	8	35.7	1.1	32.7	33.8	36.4	37.3	38.5
		High	8	151.7	1.0	144.0	150.9	152.0	153.8	156.1
	Overall	Unspiked	9	0.1	2.0	0.1	0.1	0.1	0.1	0.8
		Low #1	15	0.4	3.5	0.1	0.1	0.8	1.2	2.2
		Low #2	24	0.6	3.7	0.1	0.1	1.4	1.5	2.5
		Medium	24	36.8	1.1	32.7	35.3	36.7	38.4	41.1
		High	24	155.6	1.0	139.2	152.7	156.1	159.8	167.8
Lab B	1	Unspiked	3	5.0	1.0	5.0	5.0	5.0	5.0	5
		Low #1	5	5.0	1.0	5.0	5.0	5.0	5.0	5
		Low #2	8	13.8	4.4	5.0	5.0	5.0	46.5	196.2
		Medium	8	327.1	6.2	5.0	231.7	661.6	1052.4	1163.5
		High	8	2249.1	1.9	1062.1	1493.5	1667.8	4183.0	5751.3
Lab C	1	Unspiked	3	0.4	1.0	0.4	0.4	0.4	0.4	0.4
		Low #1	8	0.5	1.3	0.4	0.4	0.5	0.6	0.8
		Low #2	9	1.5	1.3	1.0	1.3	1.6	1.8	2.1
		Medium	10	43.7	1.3	20.2	42.0	45.0	52.1	60.7
		High	10	171.6	1.2	109.6	175.0	183.3	188.4	203
Lab D	1	Unspiked	3	1.3	1.0	1.3	1.3	1.3	1.3	1.3
		Low #1	5	1.3	1.0	1.3	1.3	1.3	1.3	1.3

		Low #2	8	1.3	1.0	1.3	1.3	1.3	1.3	1.3
		Medium	10	34.8	1.1	30.3	32.3	33.8	38.5	42.8
		High	9	105.8	1.5	38.9	97.4	125.0	128.9	137.1
All Labs	Overall	Unspiked	18	0.4	4.5	0.1	0.1	0.4	1.3	5
		Low #1	33	0.7	3.6	0.1	0.4	0.8	1.3	5
		Low #2	49	1.3	4.7	0.1	1.2	1.5	1.9	196.2
		Medium	52	52.6	2.9	5.0	35.1	38.2	44.1	1163.5
		High	51	225.3	2.9	38.9	144.0	157.4	185.6	5751.3

Note: In calculating these statistics, results reported as “below detection limits” were replaced by one-half of the laboratory’s detection limit.



**Table B-4c.Descriptive Statistics of Reported Analytical Measurements for DEP (µg/L), Calculated by Spiking Level for Each Laboratory and Across All Laboratories**

Lab	Set	Spiking Level	# Measurements	Geom. Mean	Geom. Standard Deviation	Minimum	25 <sup>th</sup> Percentile	Median	75 <sup>th</sup> Percentile	Maximum
Lab A	1	Unspiked	3	0.2	1.0	0.2	0.2	0.2	0.2	0.2
		Low #1	5	1.2	3.5	0.2	1.5	1.6	1.8	4.5
		Low #2	8	1.3	4.2	0.2	0.6	2.3	3.9	5.2
		Medium	8	46.8	1.2	40.3	40.8	44.3	52.3	65.3
		High	8	168.1	1.1	155.4	159.7	167.6	175.9	185.1
	2	Unspiked	3	0.2	1.0	0.2	0.2	0.2	0.2	0.2
		Low #1	5	0.5	4.6	0.2	0.2	0.2	2.2	2.7
		Low #2	8	0.9	4.8	0.2	0.2	1.9	3.5	4.7
		Medium	8	44.2	1.2	33.0	38.5	43.1	50.1	64.7
		High	8	168.7	1.1	155.5	157.8	169.3	177.4	188
	3	Unspiked	3	0.6	3.5	0.2	0.2	0.7	1.8	1.8
		Low #1	5	0.6	4.9	0.2	0.2	0.6	1.5	6.2
		Low #2	8	1.0	5.4	0.2	0.2	1.9	4.4	7.5
		Medium	8	42.6	1.2	35.8	36.7	39.5	49.0	61.9
		High	8	169.3	1.1	150.5	159.5	174.5	177.4	184.2
	Overall	Unspiked	9	0.2	2.5	0.2	0.2	0.2	0.2	1.8
		Low #1	15	0.7	4.1	0.2	0.2	1.5	2.2	6.2
		Low #2	24	1.1	4.5	0.2	0.2	2.1	3.6	7.5
		Medium	24	44.5	1.2	33.0	38.5	41.4	50.1	65.3
		High	24	168.7	1.1	150.5	159.1	170.5	177.1	188
Lab B	1	Unspiked	3	2.5	1.0	2.5	2.5	2.5	2.5	2.5
		Low #1	5	3.9	1.8	2.5	2.5	2.5	6.3	8.7
		Low #2	8	7.7	3.2	2.5	2.5	7.5	22.5	44.2
		Medium	8	48.6	2.8	5.4	37.7	59.6	99.4	137
		High	8	275.9	1.7	113.7	218.6	254.1	422.4	577
Lab C	1	Unspiked	3	2.1	12.3	0.5	0.5	0.5	38.7	38.7
		Low #1	8	1.8	1.6	1.0	1.2	1.6	2.8	3.5
		Low #2	9	2.9	1.8	1.3	1.7	2.9	4.7	5.5
		Medium	10	45.8	1.2	25.0	47.2	49.0	49.1	53.4
		High	10	184.3	1.2	124.5	191.6	196.5	204.0	206
Lab D	1	Unspiked	3	1.3	1.0	1.3	1.3	1.3	1.3	1.3
		Low #1	5	1.3	1.0	1.3	1.3	1.3	1.3	1.3

		Low #2	8	1.5	1.8	1.3	1.3	1.3	1.3	6.6
		Medium	10	30.8	1.2	22.2	25.7	30.9	37.0	39.5
		High	9	80.0	1.8	16.3	85.2	98.3	103.3	118.3
All Labs	Overall	Unspiked	18	0.7	4.6	0.2	0.2	0.6	1.8	38.7
		Low #1	33	1.2	3.2	0.2	1.0	1.5	2.5	8.7
		Low #2	49	1.9	4.0	0.2	1.3	2.3	4.7	44.2
		Medium	52	42.3	1.6	5.4	36.7	42.9	49.4	137
		High	51	162.5	1.6	16.3	150.5	173.8	195.9	577

Note: In calculating these statistics, results reported as “below detection limits” were replaced by one-half of the laboratory’s detection limit.

**Table B-4d. Descriptive Statistics of Reported Analytical Measurements for DEDTP ( $\mu\text{g/L}$ ), Calculated by Spiking Level for Each Laboratory and Across All Laboratories**

Lab	Set	Spiking Level	# Measurements	Geom. Mean	Geom. Standard Deviation	Minimum	25 <sup>th</sup> Percentile	Median	75 <sup>th</sup> Percentile	Maximum	
Lab A	1	Unspiked	3	0.0	3.4	0.0	0.0	0.1	0.1	0.1	
		Low #1	5	0.8	3.3	0.1	0.7	1.2	1.5	2.1	
		Low #2	8	1.3	1.9	0.3	1.2	1.5	1.9	2.5	
		Medium	8	54.0	1.1	48.1	52.0	53.5	57.2	59.5	
		High	11	162.5	2.5	22.7	224.8	239.5	250.3	283	
	2	Unspiked	3	0.1	1.0	0.1	0.1	0.1	0.1	0.1	0.1
		Low #1	5	0.3	4.3	0.1	0.1	0.1	1.0	1.9	
		Low #2	8	2.4	1.5	1.5	1.7	2.3	3.1	4.5	
		Medium	8	57.6	1.1	54.0	54.9	56.8	60.1	63.9	
		High	8	241.9	1.1	224.8	227.3	247.4	254.1	255.3	
	3	Unspiked	3	0.2	2.6	0.1	0.1	0.1	0.5	0.5	
		Low #1	5	0.3	5.5	0.1	0.1	0.1	2.0	2.6	
		Low #2	8	1.9	3.3	0.1	2.0	3.0	3.2	3.7	
		Medium	8	52.4	1.1	38.6	51.0	54.6	57.3	58.4	
		High	8	227.5	1.1	199.8	221.8	230.2	236.8	245.3	
	Overall	Unspiked	9	0.1	2.6	0.0	0.1	0.1	0.1	0.5	
		Low #1	15	0.4	4.2	0.1	0.1	0.7	1.9	2.6	
		Low #2	24	1.8	2.3	0.1	1.5	2.1	3.0	4.5	
		Medium	24	54.6	1.1	38.6	52.7	55.2	58.1	63.9	
		High	27	202.0	1.8	22.7	224.9	238.4	249.4	283	
Lab B	1	Unspiked	3	5.9	1.3	5.0	5.0	5.0	8.2	8.2	
		Low #1	5	9.7	4.4	4.9	5.0	5.0	5.0	136.9	
		Low #2	8	16.4	3.0	5.0	5.8	15.3	40.0	94.3	
		Medium	8	235.7	1.9	101.8	117.6	292.2	429.3	456.4	
		High	8	1034.6	1.7	525.3	748.9	833.6	1766.9	2066.3	
Lab C	1	Unspiked	3	0.4	1.6	0.3	0.3	0.3	0.7	0.7	
		Low #1	8	1.0	1.4	0.8	0.8	0.9	1.1	2.1	
		Low #2	9	2.4	2.4	1.1	1.5	2.2	2.3	22.9	
		Medium	10	47.5	1.3	29.4	39.8	51.8	55.0	63.8	
		High	10	201.9	1.4	118.8	152.0	240.9	246.0	281.8	
All Labs	Overall	Unspiked	15	0.3	6.2	0.0	0.1	0.1	0.7	8.2	
		Low #1	28	1.0	5.3	0.1	0.4	1.0	2.1	136.9	

		Low #2	41	3.0	3.4	0.1	1.6	2.3	3.7	94.3
		Medium	42	69.8	2.0	29.4	52.5	55.7	62.0	456.4
		High	45	270.0	2.3	22.7	225.4	242.3	263.7	2066.3

Note: In calculating these statistics, results reported as “below detection limits” were replaced by one-half of the laboratory’s detection limit.

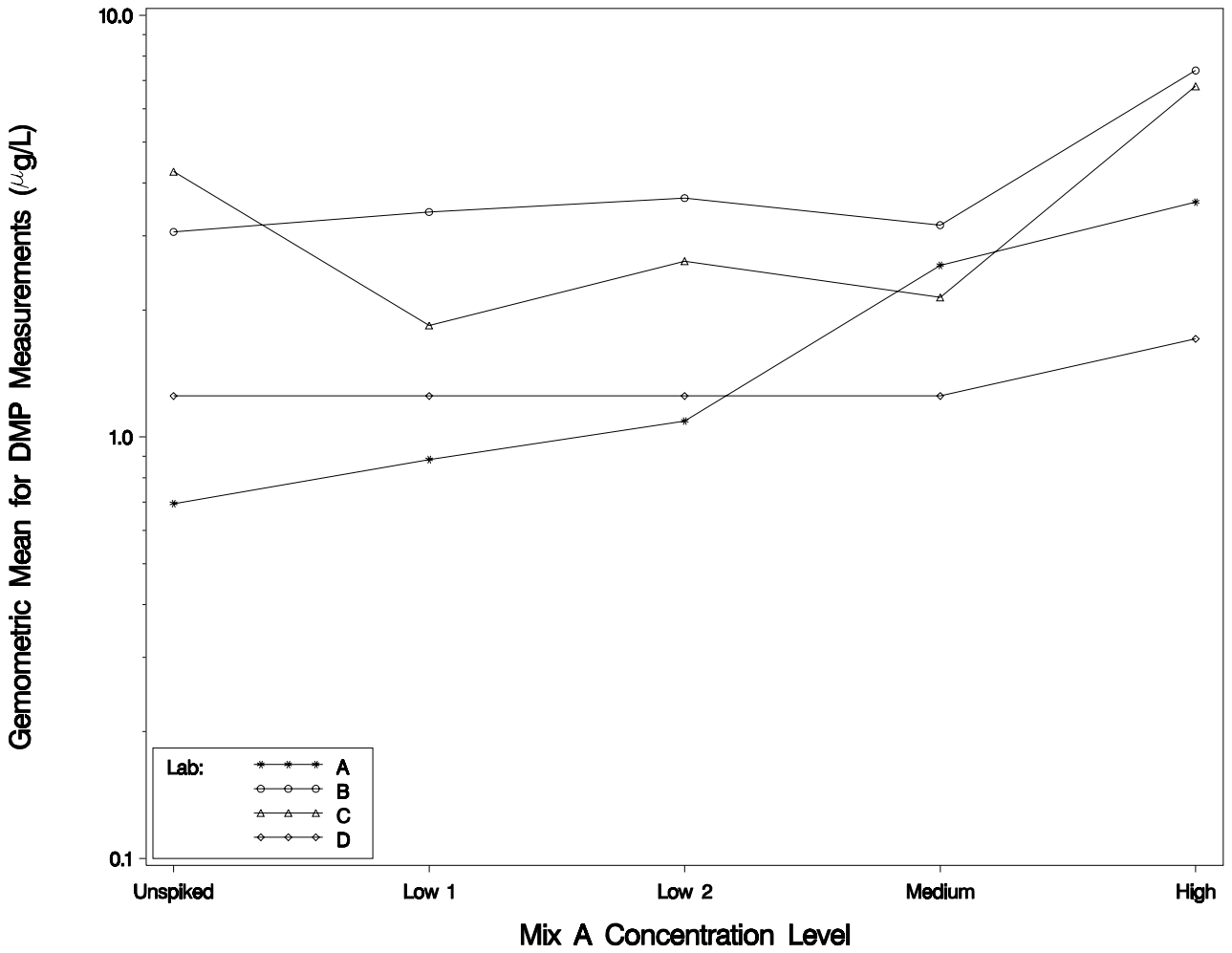
**Table B-4e. Descriptive Statistics of Reported Analytical Measurements for DMTP ( $\mu\text{g/L}$ ), Calculated by Spiking Level for Each Laboratory and Across All Laboratories**

Lab	Set	Spiking Level	# Measurements	Geom. Mean	Geom. Standard Deviation	Minimum	25 <sup>th</sup> Percentile	Median	75 <sup>th</sup> Percentile	Maximum
Lab A	1	Unspiked	3	3.0	1.0	2.9	2.9	2.9	3.0	3
		Low	13	2.2	2.2	0.4	1.8	2.0	3.1	9.2
		Medium	8	20.7	1.2	17.4	18.0	20.5	23.3	26.2
		High	8	167.9	1.1	155.1	161.9	165.9	173.2	188.1
	2	Unspiked	3	1.8	1.8	1.0	1.0	1.9	3.1	3.1
		Low	13	1.7	1.8	0.8	1.1	1.3	2.3	4.3
		Medium	8	20.0	1.2	17.1	17.2	20.1	22.5	24.8
		High	8	164.8	1.0	152.8	159.0	167.4	170.3	173.2
	3	Unspiked	3	1.6	1.4	1.1	1.1	2.0	2.0	2
		Low	13	2.6	2.4	1.0	1.1	2.2	3.8	10.3
		Medium	8	21.2	1.2	17.5	18.2	20.9	24.4	27.4
		High	8	161.8	1.0	146.3	160.7	161.9	165.8	172.2
	Overall	Unspiked	9	2.0	1.6	1.0	1.9	2.0	2.9	3.1
		Low	39	2.1	2.2	0.4	1.1	1.9	3.8	10.3
		Medium	24	20.6	1.2	17.1	17.7	20.4	23.3	27.4
		High	24	164.8	1.1	146.3	160.9	164.8	170.0	188.1
Lab B	1	Unspiked	3	6.0	4.5	2.5	2.5	2.5	34.0	34
		Low	13	7.6	3.3	2.5	2.5	8.1	19.6	57
		Medium	8	38.3	2.7	6.2	28.1	36.8	79.0	151.8
		High	8	214.6	1.8	81.5	142.5	274.0	334.4	373.2
Lab C	1	Unspiked	3	0.5	1.0	0.5	0.5	0.5	0.5	0.5
		Low	14	3.1	1.3	2.3	2.6	3.1	3.6	5
		Medium	11	20.3	1.2	14.7	16.2	21.1	23.7	26.3
		High	10	164.9	1.2	113.7	162.0	173.5	180.9	182.7
Lab D	1	Unspiked	3	1.3	1.0	1.3	1.3	1.3	1.3	1.3
		Low	13	4.2	2.7	1.3	1.3	7.0	9.7	12
		Medium	9	31.8	1.2	19.0	32.0	32.5	35.6	38.3
		High	8	157.6	1.4	75.5	156.3	176.2	181.3	202.5
All Labs	Overall	Unspiked	18	1.8	2.6	0.5	1.1	1.9	2.9	34
		Low	79	3.1	2.6	0.4	1.4	2.5	6.2	57
		Medium	52	24.4	1.6	6.2	18.8	22.0	31.6	151.8
		High	50	170.7	1.3	75.5	161.3	167.6	178.8	373.2

Note: In calculating these statistics, results reported as “below detection limits” were replaced by one-half of the laboratory’s detection limit.

**Table B-4f. Descriptive Statistics of Reported Analytical Measurements for DETP (µg/L), Calculated by Spiking Level for Each Laboratory and Across All Laboratories**

Lab	Set	Spiking Level	# Measurements	Geom. Mean	Geom. Standard Deviation	Minimum	25 <sup>th</sup> Percentile	Median	75 <sup>th</sup> Percentile	Maximum
Lab A	1	Unspiked	3	0.3	2.0	0.2	0.2	0.3	0.6	0.6
		Low	13	6.4	2.0	3.3	3.7	4.0	7.9	19.7
		Medium	8	24.4	1.2	21.3	22.0	22.8	27.2	31.9
		High	8	176.2	1.0	167.1	173.5	175.2	179.8	186.1
	2	Unspiked	3	0.3	1.7	0.2	0.2	0.4	0.4	0.4
		Low	13	5.9	1.5	4.1	4.5	4.6	6.8	10.8
		Medium	8	24.5	1.2	20.9	21.1	23.5	28.4	31.9
		High	8	181.9	1.1	165.9	176.7	183.6	187.7	195
	3	Unspiked	3	0.5	3.1	0.2	0.2	0.9	1.1	1.1
		Low	13	7.1	2.2	3.5	3.8	4.1	10.4	26.3
		Medium	8	23.6	1.2	20.0	20.3	22.0	27.6	31.7
		High	8	169.2	1.0	157.9	163.6	170.8	175.7	176.5
	Overall	Unspiked	9	0.4	2.2	0.2	0.2	0.4	0.6	1.1
		Low	39	6.4	1.9	3.3	3.9	4.5	10.3	26.3
		Medium	24	24.2	1.2	20.0	21.3	22.6	28.0	31.9
		High	24	175.7	1.1	157.9	170.5	175.7	181.9	195
Lab B	1	Unspiked	3	6.2	4.9	2.5	2.5	2.5	38.6	38.6
		Low	13	6.0	2.9	0.8	2.5	9.2	12.7	31.5
		Medium	8	33.7	2.6	4.9	25.7	35.5	63.9	121
		High	8	226.6	1.6	100.2	168.6	273.6	312.7	355.8
Lab C	1	Unspiked	3	0.6	1.5	0.5	0.5	0.5	1.0	1
		Low	14	4.2	1.4	2.7	3.2	4.1	4.4	7.4
		Medium	11	22.7	1.2	15.5	21.0	23.8	25.0	30
		High	10	182.0	1.1	137.2	182.4	186.0	191.1	195.9
Lab D	1	Unspiked	3	1.3	1.0	1.3	1.3	1.3	1.3	1.3
		Low	13	5.0	2.0	1.3	4.4	5.7	7.2	12
		Medium	9	32.0	1.2	23.9	30.8	31.7	33.0	41.8
		High	8	161.2	1.3	96.4	155.6	167.1	188.3	198
All Labs	Overall	Unspiked	18	0.8	3.8	0.2	0.4	0.8	1.3	38.6
		Low	79	5.7	2.0	0.8	3.7	4.5	9.2	31.5
		Medium	52	26.3	1.5	4.9	22.0	24.3	31.7	121
		High	50	181.8	1.3	96.4	168.2	179.3	190.0	355.8



Note: In calculating these statistics, results reported as “below detection limits” were replaced by one-half of the laboratory’s detection limit.

Figure B-1. Geometric Means of DMP Measurements (µg/L) at Each Mix A Spiking Level, Calculated for Each Laboratory

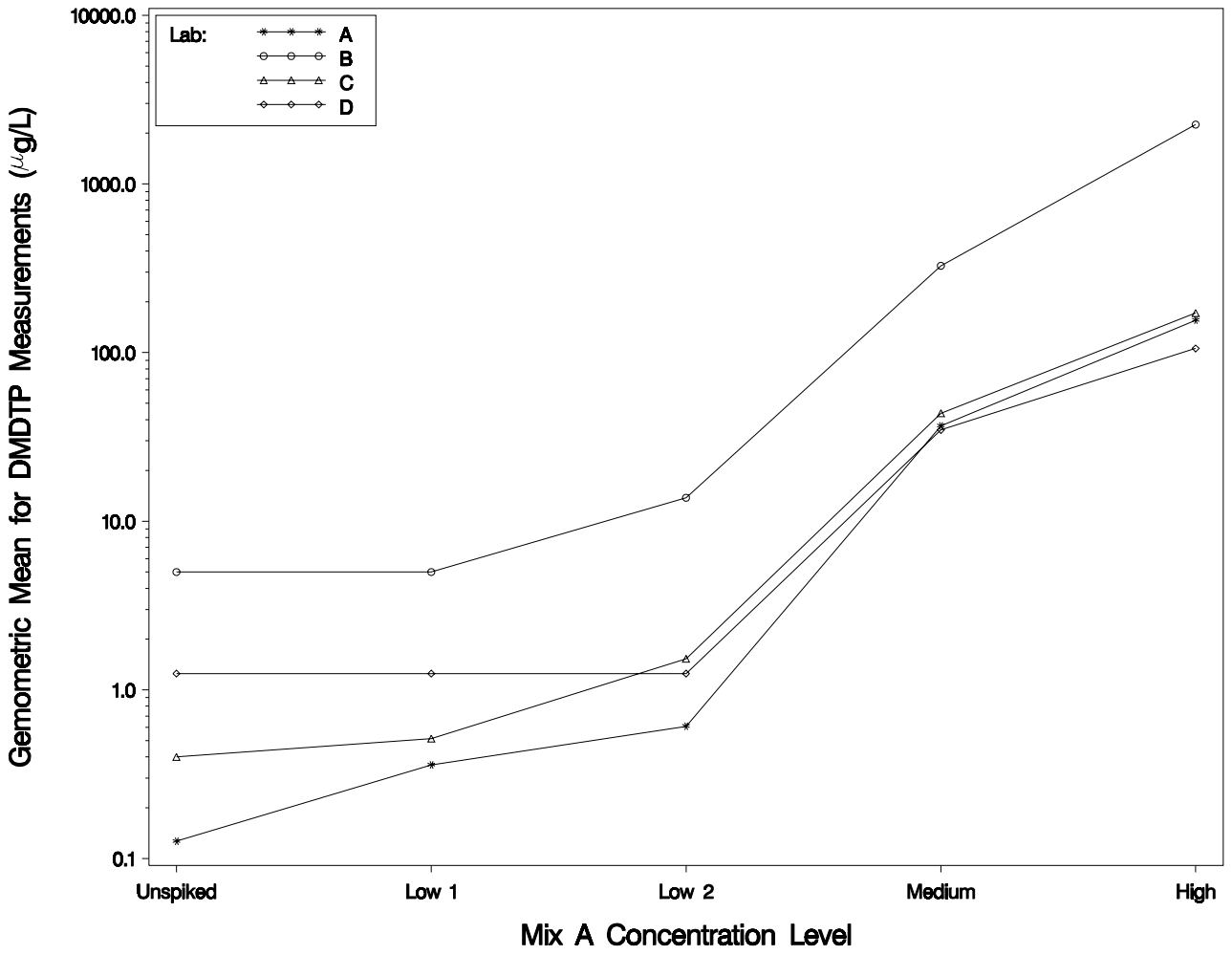


Figure B-2. Geometric Means of DMDTP Measurements (µg/L) at Each Mix A Spiking Level, Calculated for Each Laboratory



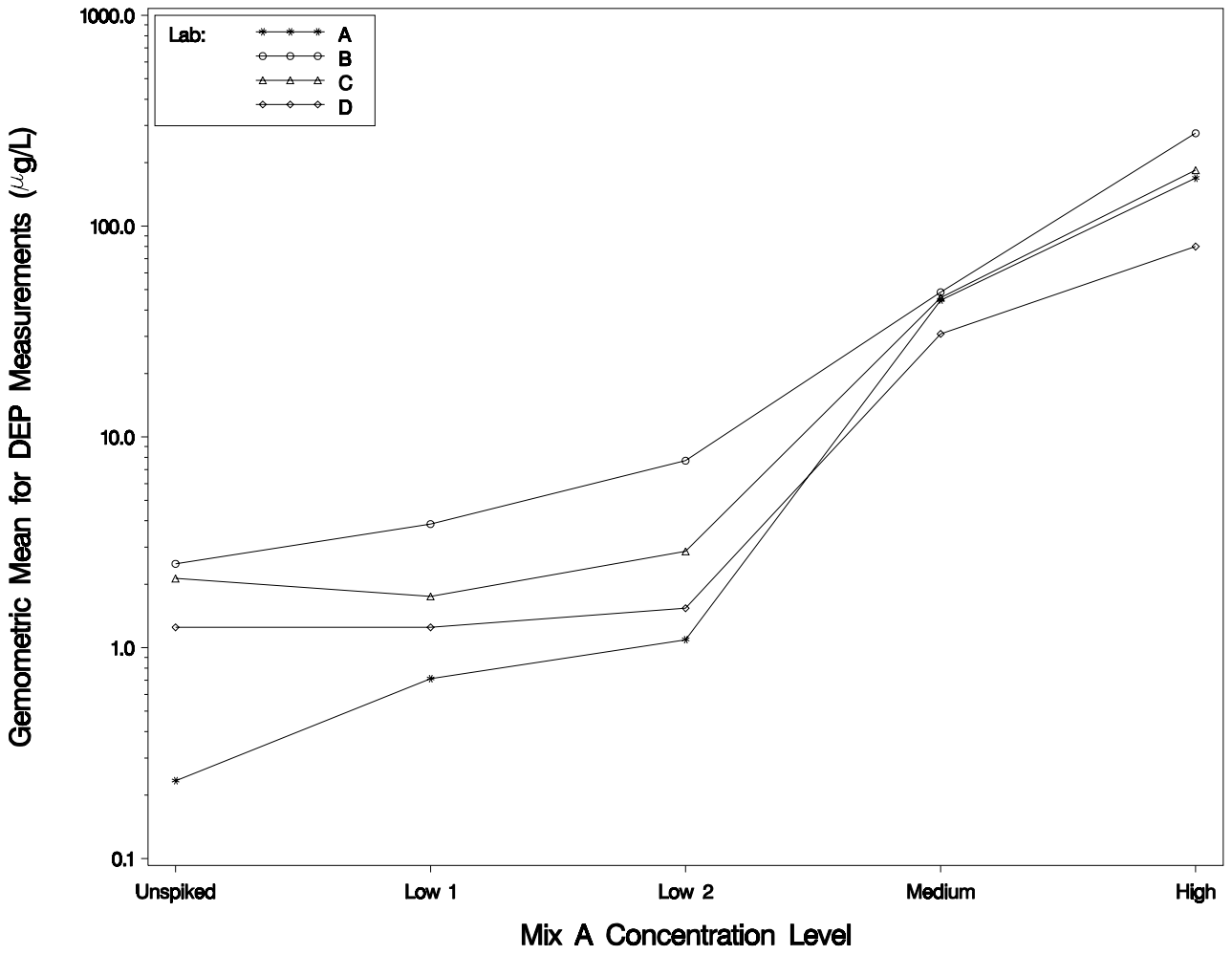


Figure B-3. Geometric Means of DEP Measurements ( $\mu\text{g/L}$ ) at Each Mix A Spiking Level, Calculated for Each Laboratory

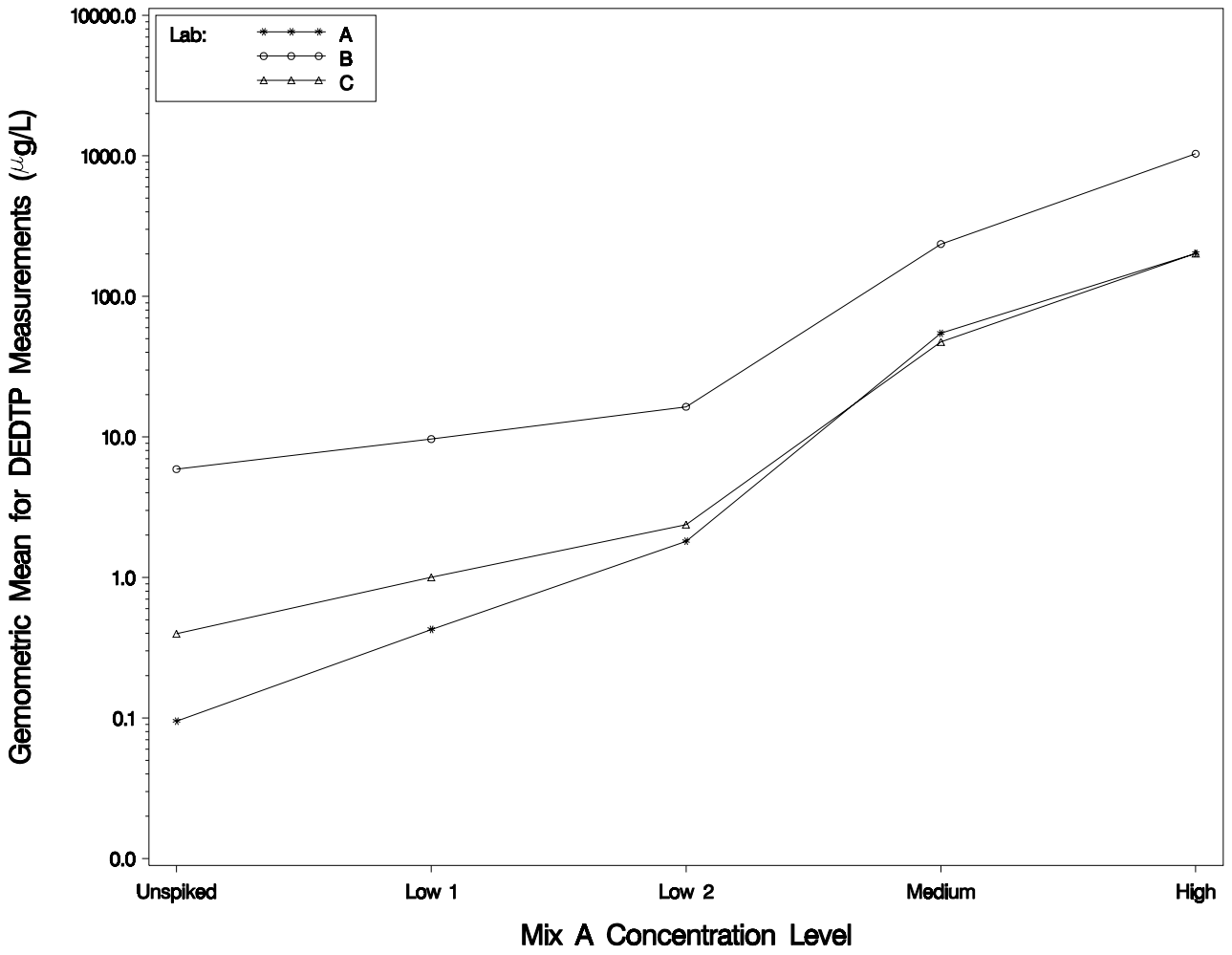


Figure B-4. Geometric Means of DEDTP Measurements ( $\mu\text{g/L}$ ) at Each Mix A Spiking Level, Calculated for Each Laboratory

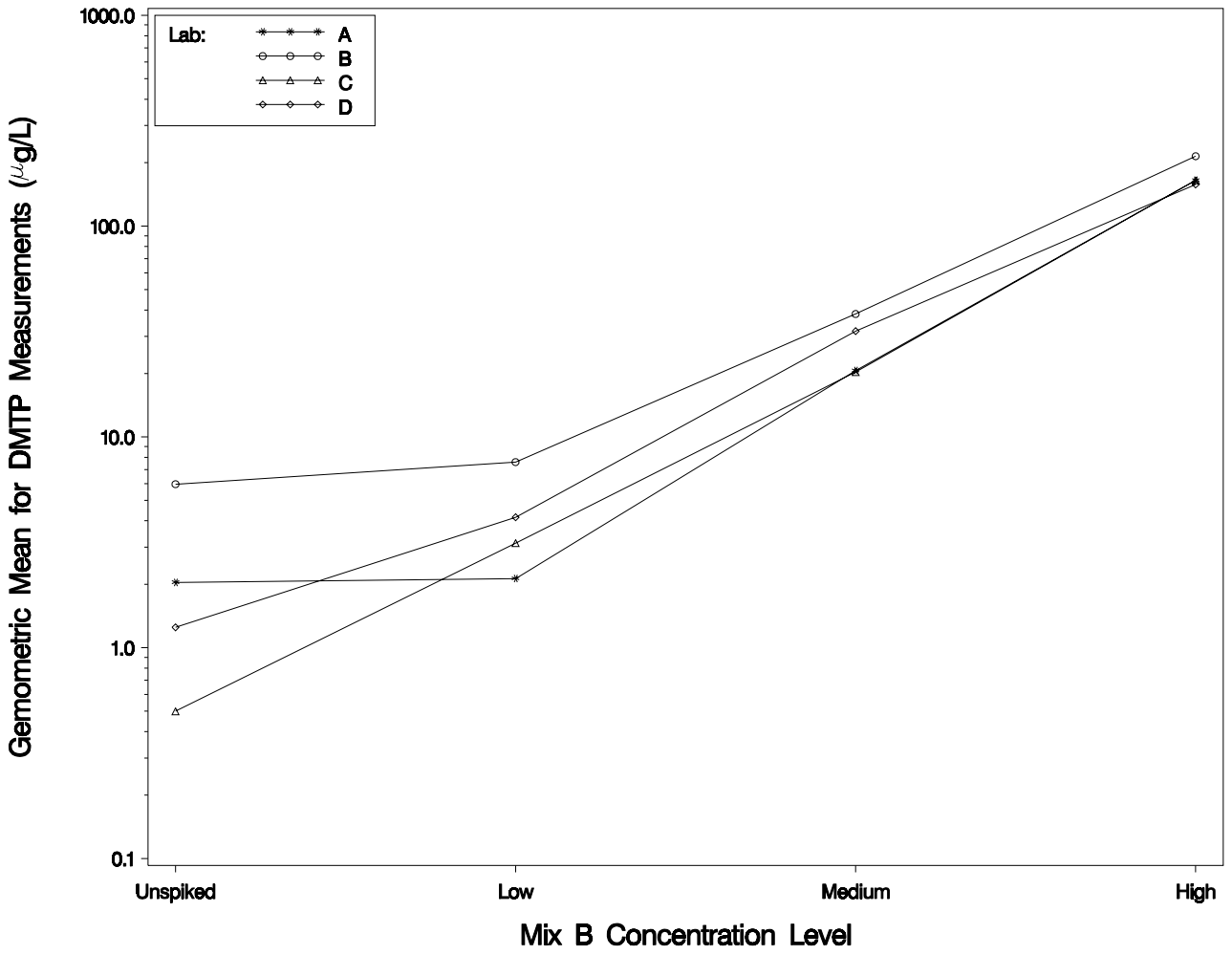


Figure B-5. Geometric Means of DMTP Measurements ( $\mu\text{g/L}$ ) at Each Mix B Spiking Level, Calculated for Each Laboratory

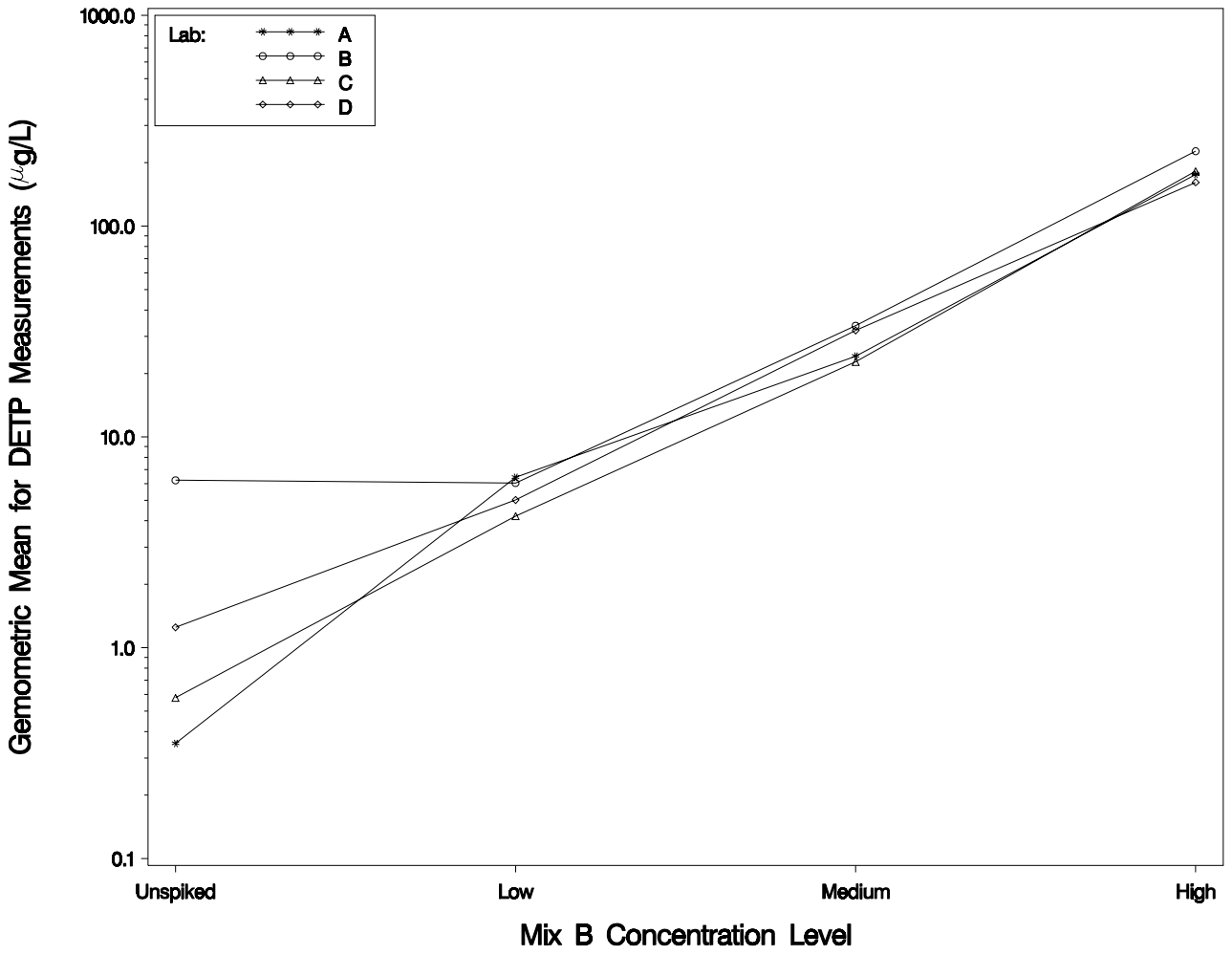


Figure B-6. Geometric Means of DETP Measurements ( $\mu\text{g/L}$ ) at Each Mix B Spiking Level, Calculated for Each Laboratory

Results of the statistical analyses, involving fitting ANOVA model (1) to the log-transformed measurements, are summarized in Tables B-5 through B-7. Table B-5 contains the p-values for the tests of fixed effects (i.e., lab effects, spiking level effects, and their interactions) included in the model. The results of statistical tests to further investigate the presence of significant differences among spiking levels for each laboratory, as well as overall across all laboratories, are presented in Table B-6. Table B-7 contains the results of statistical tests for differences among the laboratories at each spiking level, as well as overall across all spiking levels. Selected findings from these tables are found in Tables 6 and 10 of the main report.

**Table B-5. Summary of Tests for Fixed Effects Included in the ANOVA Model, For each Alkyl Phosphate Target Compound**

Fixed Effects	Results of Statistical Test for Fixed Effect (p-values)					
	Mix A				Mix B	
	DMP	DMDTP	DEP	DEDTP	DMTP	DETP
Lab	0.0527	0.0054	<0.0001	<0.0001	0.0214	0.0265
Mix A Spiking <sup>a</sup>	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
Mix B Spiking <sup>b</sup>	<0.0001	0.6056	0.0049	0.9921	<0.0001	<0.0001
(Mix A Spiking)*(Mix B Spiking)	0.2492	0.577	0.0063	0.3527	<0.0001	<0.0001
Lab*(Mix A Spiking) <sup>a</sup>	0.0028	0.0069	0.0005	0.0001	<0.0001	<0.0001
Lab*(Mix B Spiking) <sup>b</sup>	<0.0001	0.5122	0.1910	0.2334	<0.0001	<0.0001
Lab*(Mix A Spiking)* (Mix B Spiking)	0.1410	0.4290	0.3511	0.0146	<0.0001	<0.0001

a. For compounds in Mix A, these are the spiking level effects and the interactions of laboratory and spiking levels.

b. For compounds in Mix B, these are the spiking level effects and the interactions of laboratory and spiking levels.

**Table B-6. Summary of Spiking Level Effects for Each Lab and Overall across all Labs for each Alkyl Phosphate Target Compound<sup>a,b,c</sup>**

Lab	Significant Spiking Level Effect?					
	Mix A				Mix B	
	DMP	DMDTP	DEP	DEDTP	DMTP	DETP
Lab A	Yes (H vs. U, L1, L2; M vs. U, L1, L2)	Yes (H vs. U, L1, L2, M; M vs. U, L1, L2; L2 vs. U; L1 vs. U)	Yes (H vs. U, L1, L2, M; M vs. U, L1, L2; L2 vs. U; L1 vs. U;)	Yes (H vs. U, L1, L2, M; M vs. U, L1, L2; L2 vs. U, L1; L1 vs. U)	Yes (H vs. U, L, M; M vs. U, L)	Yes (H vs. U, L, M; M vs. U, L; L vs. U)
Lab B	No	Yes (H vs. U, L1, L2, M; M vs. U, L1, L2)	Yes (H vs. U, L1, L2, M; M vs. U, L1, L2)	Yes (H vs. U, L1, L2, M; M vs. U, L1, L2;)	Yes (H vs. U, L, M; M vs. U, L)	Yes (H vs. U, L, M; M vs. U, L)
Lab C	Yes (H vs. L1, L2, M)	Yes (H vs. U, L1, L2, M; M vs. U, L1, L2)	Yes (H vs. U, L1, L2, M; M vs. U, L1, L2)	Yes (H vs. U, L1, L2, M; M vs. U, L1, L2; L2 vs. U)	Yes (H vs. U, L, M; M vs. U, L; L vs. U)	Yes (H vs. U, L, M; M vs. U, L; L vs. U)
Lab D	No	Yes (H vs. U, L1, L2, M vs. U, L1, L2)	Yes (H vs. U, L1, L2; M vs. U, L1, L2)	---	Yes (H vs. U, L, M; M vs. U, L; L vs. U)	Yes (H vs. U, L, M; M vs. U, L; L vs. U)
All Labs	Yes (H vs. U, L1, L2, M; M vs. U, L1)	Yes (H vs. U, L1, L2, M; M vs. U, L1, L2; L2 vs. L1, U)	Yes (H vs. U, L1, L2, M; M vs. U, L1, L2; L2 vs. U)	Yes (H vs. U, L1, L2, M; M vs. U, L1, L2; L2 vs. U, L1; L1 vs. U)	Yes (H vs. U, L, M; M vs. U, L; L vs. U)	Yes (H vs. U, L, M; M vs. U, L; L vs. U)

- F tests were used to test for significant lab effects across all labs at the 0.05 level. When significant differences among the labs were present, pairwise comparisons were made between each pair of spiking level, with each pairwise comparison performed using Bonferroni-adjustment method to ensure that the overall error rate across the pairwise comparisons was no greater than 0.05. Pairs of spiking levels differing significantly at the Bonferroni-adjusted 0.05 level are identified in parentheses.
- F tests were used to test for significant spiking level effects for each lab, where the Benjamini and Hochberg multiple comparison adjustment method was used to control the overall error rate across all of these tests to be no higher than 0.05. When significant differences among spiking levels were present for a given lab, pairwise comparisons were made between each pair of spiking levels for the given lab, with each pairwise comparison performed using Bonferroni-adjustment method to ensure that the overall error rate across the pairwise comparisons was no greater than 0.05. Pairs of spiking levels differing significantly at the Bonferroni-adjusted 0.05 level are identified in parentheses.
- Mix A compounds and Mix B compounds were spiked at five and four spiking levels, respectively.

**Table B-7. Summary of Lab Effects at each Spiking Level and Overall across all Spiking Levels for each Alkyl Phosphate Target Compound <sup>a,b</sup>**

Spiking Level <sup>c</sup>	Significant Lab Effect?					
	Mix A				Mix B	
	DMP	DMDTP	DEP	DEDTP	DMTP	DETP
Unspiked	Yes (Lab A vs. Lab B, Lab C)	Yes (Lab B vs Lab A, Lab C; Lab A vs Lab D)	Yes (Lab A vs Lab B, Lab C, Lab D)	Yes (Lab B vs Lab A, Lab C; Lab A vs Lab C)	Yes (Lab B vs Lab A, Lab C, Lab D; Lab A vs Lab C)	Yes (Lab A vs Lab B, Lab C, Lab D; Lab B vs Lab C, Lab D; Lab C vs Lab D)
Low #1	Yes (Lab A vs Lab B)	Yes (Lab B vs Lab A, Lab C; Lab A vs Lab D)	Yes (Lab A vs Lab B)	Yes (Lab B vs Lab A, Lab C; Lab A vs Lab C)	Yes (Lab A vs Lab B, Lab D; Lab B vs Lab C)	Yes (Lab A vs Lab C)
Low #2	Yes (Lab A vs Lab B)	Yes (Lab B vs Lab A, Lab C, Lab D)	Yes (Lab B vs Lab A, Lab D; Lab A vs Lab C)	Yes (Lab B vs Lab A, Lab C)		
Medium	No	Yes (Lab B vs Lab A, Lab C, Lab D)	No	Yes (Lab B vs Lab A, Lab C)	Yes (Lab A vs Lab B)	Yes
High	Yes (Lab D vs Lab B, Lab C)	Yes (Lab B vs Lab A, Lab C, Lab D)	Yes (Lab B vs Lab D)	Yes (Lab B vs Lab A, Lab C)	No	No
Overall	No	Yes (Lab A vs Lab B)	Yes (Lab B vs Lab A, Lab C, Lab D; Lab C vs Lab A)	Yes (Lab B vs Lab A, Lab C)	Yes	Yes

- F tests were used to test for significant lab effects across all spiking levels at the 0.05 level. When significant differences among the labs were present, pairwise comparisons were made between each pair of labs, with each pairwise comparison performed using Bonferroni-adjustment method to ensure that the overall error rate across the pairwise comparisons was no greater than 0.05. Pairs of labs differing significantly at the Bonferroni-adjusted 0.05 level are identified in parentheses.
- F tests were used to test for significant lab effects at each spiking level of the compound, where the Benjamini and Hochberg multiple comparison adjustment method was used to control the overall error rate across all of these tests to be no higher than 0.05. When significant differences among labs were present at a given spiking level, pairwise comparisons were made between each pair of labs at the given spiking level, with each pairwise comparison performed using Bonferroni-adjustment method to ensure that the overall error rate across the pairwise comparisons was no greater than 0.05. Pairs of labs differing significantly at the Bonferroni-adjusted 0.05 level are identified in parentheses.
- Mix A compounds and Mix B compounds were spiked at five and four spiking levels, respectively.

Based upon adjusted significance levels, the spiking level effect was found to be significant for each laboratory and each compound, except for DMP at Lab D. Thus, pairwise comparisons were made in each of these instances to identify those pairs of spiking levels that differed significantly from each other for a given laboratory and compound. The findings of these tests, presented within Table B-6, were as follows:

- a. For DMP, the High and Medium levels were significantly different from (and higher than) the Unspiked, Low #1, and Low #2 levels, for Lab A. For Lab C, the High spiking level was significantly different from (and higher than) the Low #1, Low #2, and Medium levels. In particular, the Unspiked level was not significantly different from any other spiking level among Labs B, C, and D, and was significantly different only from the High level for Lab A.
- b. For DMDTP, the High and Medium spiking levels were significantly different from (and higher than) the Unspiked, Low #1, and Low #2 levels for each laboratory. Also, for Labs A, B, and C, the High spiking level was significantly different from (and higher than) the Medium level, and for Lab A, the Unspiked level was significantly different from (and lower than) the Low #1 and Low #2 levels.
- c. For DEP, the High and Medium spiking levels were significantly different from (and higher than) the Unspiked, Low #1, and Low #2 levels for each laboratory. Also, for Labs A, B, and C, the High spiking level was significantly different from (and higher than) the Medium level, and for Lab A, the Unspiked level was significantly different from (and lower than) the Low #1 and Low #2 levels.
- d. For DEDTP (analyzed by only Labs A, B, and C), for each laboratory, the High and Medium spiking levels were significantly different from (and higher than) the Unspiked, Low #1, and Low #2. The High spiking level was significantly different from (and higher than) the Medium level for Labs A, B, and C. Also for Lab A, the Unspiked level was significantly different from (and lower than) the Low #1 and Low #2 levels, and for Lab C, the Unspiked level was significantly different from (and lower than) the Low #2 level.
- e. For DMTP, each of the spiking levels was significantly different from each of the other levels for all laboratories except Lab A and Lab B, when the Low and Unspiked levels were not significantly different.
- f. For DETP, each of the spiking levels was significantly different from each of the other levels for all laboratories except Lab B, when the Low and Unspiked levels were not significantly different.

Also based upon adjusted significance levels, the laboratory effect was found to be significant at certain spiking levels for certain compounds. In these instances, pairwise comparisons



were made to identify those pairs of laboratories that differed significantly from each other. The findings of these tests, presented within Table B-7, were as follows:

- 1) For DMP, laboratory effects were significant at all spiking levels except the medium level. Lab A was significantly different from (and lower than) Lab B at the Unspiked, Low #1 and Low #2 levels, and Lab C at the Unspiked level. At the high level, Lab D was significantly different from (and lower than) the other three laboratories.
- 2) For DMDTP, laboratory effects were significant at all five spiking levels. Lab B was significantly different from (and higher than) Labs A and C at all five spiking levels and Lab D at Low #2, Medium, and High levels. In addition, Lab A was significantly different from (and lower than) Lab D at the Unspiked and Low #1 levels.
- 3) For DEP, laboratory effects were significant at the Unspiked, Low #1, Low #2, and High levels. Lab B was significantly different from (and higher than) Lab A at the Unspiked, Low #1, and Low #2 levels, and Lab D at the Low #2, and High levels. In addition, Lab A was significantly different from (and lower than) Lab C at the Unspiked and Low #2 levels, and Lab D at the Unspiked level.
- 4) For DEDTP, laboratory effects were significant at all five spiking levels. Lab B was significantly different from (and higher than) Labs A and C at all five levels. In addition, Lab A was significantly different from (and lower than) Lab C at the Unspiked and Low #1 levels..
- 5) For DMTP, laboratory effects were significant at the Unspiked, Low, and Medium levels. Lab B was significantly different from (and higher than) Lab A at each of these three levels, Lab C at the Unspiked and Low levels, and from Lab D at the Unspiked level. In addition, Lab A was significantly different from (and lower than) Lab D at the Low level; Lab A was significantly different from (and higher than) Lab C at the Unspiked level; and Lab C was significantly different from (and lower than) Lab D at the Unspiked level.
- 6) For DETP, laboratory effects were significant at the Unspiked, Low, and Medium Level; At the Unspiked Level, all labs were found to be significantly different from each other with the highest value in Lab C, following by Lab B, then by Labs D and A. In addition, Lab A was significantly different from (and higher than) the Lab C at the Low level. No significant differences between any two labs were observed at the Medium level.

## Appendix C Raw Data

### Lab A Set 1

Sample Number	Reported Measurement (ug/L)					
	DMP	DMDTP	DEP	DEDTP	DMTP	DETP
1	4.22	161.41	158.08	239.49	5.79	11.52
2	0.51	<0.2	3.03	2.52	1.38	3.27
3	1.15	<0.2	0.67	0.02	1.19	0.56
4	14.74	158.25	177.91	263.67	188.08	174.52
5	0.56	<0.2	2.61	1.88	2.94	0.60
6	10.19	0.38	23.22	0.39	167.00	186.10
7	0.65	41.08	44.30	54.15	2.01	7.88
8	0.52	<0.2	1.76	0.75	17.61	21.35
9	2.93	35.80	44.23	51.45	21.00	23.62
10	7.76	0.87	4.81	1.21	163.31	172.79
11	1.39	<0.2	<0.3	<0.2	1.83	4.04
12	1.66	1.71	1.94	1.46	17.40	21.94
13	0.35	154.39	169.77	28.48	6.24	19.71
13	.	.	.	22.73	.	.
13	.	.	.	250.29	.	.
13	.	.	.	224.81	.	.
14	12.99	36.54	65.30	59.51	170.09	175.85
15	<0.6	1.01	1.46	1.54	0.39	3.80
16	1.74	35.15	41.09	59.48	3.07	7.73
17	5.99	157.65	161.40	229.26	25.57	30.75
18	1.50	2.49	<0.3	1.57	1.87	4.01
19	<0.6	1.21	1.58	1.20	1.97	3.65
20	1.94	<0.2	1.03	0.33	18.37	22.07
21	0.97	<0.2	<0.3	<0.2	3.01	0.26
22	1.60	160.59	173.80	283.04	9.24	18.84
23	2.03	<0.2	4.49	2.08	160.51	174.11
24	2.60	0.34	3.48	0.73	20.07	22.07
25	4.47	155.69	165.49	243.45	26.24	31.87
26	<0.6	35.66	40.33	52.94	2.07	7.77
27	14.23	162.39	185.15	239.73	176.22	182.87
28	1.50	157.36	155.44	238.90	5.15	17.73
29	2.55	38.31	44.43	54.89	2.18	3.04
30	15.81	35.13	60.08	52.46	164.89	176.67
31	1.74	<0.2	5.17	1.97	155.15	167.14
32	3.25	35.75	40.51	48.06	20.96	23.66
33	<0.6	<0.2	<0.3	0.01	1.75	3.65
34	<0.6	1.89	<0.3	1.15	1.21	3.68
35	1.65	<0.2	<0.3	<0.2	2.92	<0.3

### Lab A Set 2

Sample Number	Reported Measurement (ug/L)					
	DMP	DMDTP	DEP	DEDTP	DMTP	DETP
1	3.82	156.80	175.18	224.82	5.63	10.09
2	<0.6	0.99	2.22	2.29	1.04	4.46
3	1.23	<0.2	0.79	0.46	1.56	0.43
4	13.53	156.10	167.24	228.15	170.69	186.32
5	<0.6	1.49	1.60	1.52	3.07	0.40
6	9.25	.	16.15	<0.2	161.73	187.27
7	<0.6	40.03	38.30	57.33	2.15	6.44
8	1.51	0.77	<0.3	<0.2	17.11	21.09
9	2.98	38.02	48.20	54.30	20.09	24.10
10	1.83	1.35	4.74	3.16	169.92	180.97
11	0.97	<0.2	<0.3	<0.2	1.33	4.48
12	3.18	1.44	2.70	2.41	17.21	21.17
13	<0.6	159.35	156.57	245.34	4.26	10.27
14	13.80	38.05	64.67	53.97	166.61	177.70
15	1.14	0.81	<0.3	1.02	1.06	4.58
16	2.58	35.48	46.44	58.10	2.29	6.81
17	5.30	163.31	171.30	249.38	24.12	31.89
18	1.07	1.46	<0.3	2.96	1.17	4.54
19	1.74	<0.2	2.22	1.86	0.89	4.28
20	1.71	<0.2	<0.3	4.52	17.13	20.92
21	2.31	<0.2	<0.3	<0.2	1.92	0.36
22	2.54	153.09	188.02	226.44	4.02	10.81
23	1.24	1.23	2.72	<0.2	173.23	188.18
24	2.27	1.56	4.82	0.36	20.07	22.85
25	6.27	160.33	159.11	255.27	24.78	31.39
26	1.52	40.19	39.66	63.86	1.91	5.72
27	13.12	167.79	179.61	254.09	168.18	195.03
28	1.37	139.20	155.51	254.06	4.09	10.64
29	1.78	39.68	38.66	56.23	2.78	4.39
30	10.44	39.38	52.10	62.02	156.21	175.64
31	4.82	1.15	4.23	1.55	152.76	165.91
31	5.73	.	.	.	.	.
32	3.41	33.48	32.97	55.51	20.83	25.35
32	2.59	.	.	.	.	.
33	1.34	0.83	<0.3	<0.2	0.85	4.14
34	<0.6	1.52	<0.3	1.88	1.13	4.33
35	1.55	<0.2	<0.3	<0.2	0.97	<0.3

**Lab A Set 3**

Sample Number	Reported Measurement (ug/L)					
	DMP	DMDTP	DEP	DEDTP	DMTP	DETP
1	4.46	151.76	174.12	218.65	6.72	10.43
2	<0.6	1.54	<0.3	3.34	0.98	3.80
3	0.81	0.46	0.54	<0.2	2.19	1.05
4	17.48	150.17	176.36	224.88	172.25	176.49
5	<0.6	1.50	1.51	2.98	1.95	0.95
6	12.80	0.49	21.79	1.07	161.53	164.65
7	<0.6	37.78	36.43	58.15	3.80	10.44
8	0.54	1.37	1.46	<0.2	17.47	20.02
9	3.79	38.52	41.49	52.50	22.39	24.30
10	2.56	<0.2	7.47	3.16	160.09	162.50
11	<0.6	<0.2	<0.3	<0.2	1.08	3.89
12	2.05	1.48	2.34	<0.2	17.85	20.07
13	2.26	156.12	174.79	199.83	9.29	22.09
14	15.06	35.94	61.93	58.43	166.89	175.90
15	<0.6	<0.2	<0.3	<0.2	1.11	3.81
16	1.71	33.11	36.94	49.41	3.01	9.06
17	6.28	151.70	150.47	238.38	26.47	30.93
18	<0.6	2.07	<0.3	3.09	1.10	4.07
19	<0.6	<0.2	<0.3	<0.2	1.17	3.74
20	<0.6	<0.2	<0.3	3.73	18.65	21.63
20	1.89	.	.	.	.	.
21	1.59	0.76	0.55	1.96	1.07	1.13
22	1.49	154.14	159.16	225.40	10.28	26.32
23	2.53	2.19	6.16	2.62	164.80	173.42
24	2.73	0.65	2.89	<0.2	20.75	20.50
25	6.14	153.45	178.50	235.31	27.39	31.75
26	<0.6	36.81	35.81	53.25	3.38	7.96
27	16.47	152.26	184.20	235.06	162.29	175.55
28	<0.6	143.99	159.84	245.32	10.09	21.81
29	2.78	36.87	37.76	55.98	2.70	3.11
30	13.56	34.48	56.60	56.38	161.33	168.20
31	3.02	1.83	6.47	1.79	146.29	157.86
32	3.36	32.71	41.24	38.57	21.09	22.40
33	<0.6	<0.2	0.66	0.52	1.82	3.52
34	0.87	<0.2	2.42	2.29	2.23	3.95
35	<0.6	<0.2	1.83	<0.2	1.98	<0.3

**Lab B**

Sample Number	Reported Measurement (ug/L)					
	DMP	DMDTP	DEP	DEDTP	DMTP	DETP
1	10.40	1529.40	232.30	757.70	130.80	102.20
2	<5	<10	<5	<10	<5	<5
3	<5	<10	3.40	6.50	<5	1.90
4	11.80	1535.30	315.80	810.70	373.20	355.80
5	13.30	196.20	44.20	94.30	34.00	38.60
6	3.90	145.40	13.60	331.80	81.50	100.20
7	<5	1066.60	51.00	456.40	8.10	10.40
8	<5	<10	6.30	136.90	41.30	39.20
9	<5	268.00	28.40	130.90	22.20	22.10
10	3.30	<10	10.20	36.70	229.50	286.50
11	4.60	<10	<5	<10	8.10	9.20
12	3.10	40.50	34.70	20.20	36.00	29.20
13	4.70	1062.10	113.70	525.30	22.80	31.50
14	8.70	195.40	70.90	101.80	188.80	208.00
15	<5	<10	<5	4.90	<5	3.10
16	<5	1038.10	127.80	419.60	16.60	9.90
17	11.20	1800.20	236.80	856.40	151.80	121.00
18	<5	<10	<5	6.60	<5	0.80
19	<5	<10	<5	<10	<5	<5
20	<5	<10	5.40	10.40	37.60	37.90
21	<5	<10	<5	<10	<5	<5
22	6.10	4497.90	529.00	1895.80	41.60	17.90
23	11.90	<10	8.70	<10	348.90	338.90
24	<5	<10	4.20	10.30	33.90	33.00
25	8.20	1457.50	204.90	740.10	116.70	88.50
26	<5	1163.50	137.00	439.00	19.60	12.70
27	11.10	3868.10	271.40	1638.00	318.40	274.70
28	<5	5751.30	577.00	2066.30	57.00	18.90
29	<5	882.60	68.20	353.90	15.10	5.20
30	4.90	440.60	47.00	230.50	96.10	129.10
31	6.40	52.50	9.50	43.30	319.90	272.50
32	<5	<10	5.40	104.20	6.20	4.90
33	<5	<10	<5	8.20	<5	2.90
34	<5	<10	<5	<10	<5	<5
35	<5	<10	<5	<10	<5	<5

**Lab C**

Sample Number	Reported Measurement (ug/L)					
	DMP	DMDTP	DEP	DEDTP	DMTP	DETP
1	6.9	196.2	199.6	152.0	3.3	5.6
1	7.0	203.0	204.0	149.0	5.0	9.0
2*	<1.6	1.0	1.7	1.1	2.3	3.0
3	3.7	<0.8	1.9	<0.6	<1	1.5
4	10.3	138.9	146.3	118.8	113.7	137.2
5*	3.3	1.5	5.5	1.2	<1	<0.6
6	10.8	<0.8	8.6	<0.6	165.2	188.3
7*	3.6	20.2	25.0	29.4	3.0	3.3
8*	<1.6	<0.8	1.6	0.9	14.7	15.5
8	<1.6	<0.8	1.0	0.9	15.0	16.0
9	<1.6	60.7	49.7	39.7	21.1	23.1
9	2.0	56.0	47.0	42.0	23.0	25.0
10	5.9	1.2	4.7	2.2	180.9	183.7
11	120.6	<0.8	38.7	0.7	3.5	4.2
12	2.3	1.8	2.9	2.3	21.9	24.8
13*	6.8	109.6	124.5	152.3	3.6	7.4
14	5.9	52.1	53.4	39.8	152.0	195.9
15	<1.6	<0.8	1.5	1.3	3.3	4.3
15	<1.6	<0.8	1.0	0.8	3.0	3.0
16	2.5	42.0	49.0	63.8	3.6	4.1
17	8.3	185.6	204.4	281.8	26.3	27.0
17	6.0	181.0	206.0	243.0	26.0	30.0
18*	2.1	1.3	1.7	22.9	2.6	3.2
19	5.7	0.8	2.6	2.1	3.0	4.0
20	4.2	1.8	2.1	2.1	21.1	23.7
21*	2.9	<0.8	1.3	0.8	<1	1.0
22	7.7	177.1	195.9	281.0	5.0	7.3
23	4.2	0.8	3.5	0.9	181.0	191.1
23	3.0	<0.8	3.0	0.8	176.0	190.0
24	<1.6	0.9	1.7	<0.6	23.7	23.8
25	5.1	188.4	191.6	246.0	18.9	21.0
26	1.7	43.5	49.1	59.1	3.2	4.4
27	9.9	185.9	197.0	242.3	182.7	194.3
28	3.2	175.0	195.5	239.4	4.2	7.3
29	<1.6	46.3	48.9	53.9	1.1	5.1
29	<1.6	41.0	48.0	55.0	2.0	5.0
30	7.3	45.3	49.1	53.3	178.8	182.4
31	4.5	1.6	4.6	2.3	171.0	183.1
31	5.0	2.0	5.0	2.2	162.0	182.0
32	2.5	44.7	47.2	50.2	16.2	24.2
33	<1.6	<0.8	<1	<0.6	2.3	2.7
34	<1.6	2.1	1.3	1.5	2.4	4.2
35	<1.6	<0.8	<1	<0.6	<1	<0.6

\* Sample was received at laboratory in a broken vial, but samples was still frozen so the analysis was completed. No contamination problems were observed.

**Lab D**

Sample Number	Reported Measurement (ug/L)					
	DMP	DMDTP	DEP	DEDTP	DMTP	DETP
1	<2.5	134.60	118.30	.	7.30	9.20
1	<2.5	128.90	99.00	.	5.70	9.30
2	<2.5	<2.5	<2.5	.	<2.5	<2.5
3	<2.5	<2.5	<2.5	.	7.90	<2.5
4	<2.5	120.40	98.30	.	183.90	185.60
5	<2.5	<2.5	<2.5	.	<2.5	<2.5
6	<2.5	<2.5	<2.5	.	174.50	163.70
7	<2.5	30.30	25.70	.	<2.5	4.40
8	<2.5	<2.5	<2.5	.	19.00	23.90
9	<2.5	33.40	37.00	.	32.20	30.50
10	<2.5	<2.5	<2.5	.	75.50	96.40
11	<2.5	<2.5	<2.5	.	7.00	5.50
12	<2.5	<2.5	<2.5	.	36.60	31.60
13	5.90	128.70	103.30	.	12.00	12.00
14	<2.5	34.30	39.50	.	177.90	190.90
15	<2.5	<2.5	<2.5	.	<2.5	3.60
16	<2.5	42.80	32.10	.	8.70	9.30
17	<2.5	137.10	92.40	.	31.20	30.80
18	<2.5	<2.5	<2.5	.	6.20	5.70
19	<2.5	<2.5	<2.5	.	9.70	6.70
20	<2.5	<2.5	<2.5	.	35.60	31.70
21	<2.5	<2.5	<2.5	.	<2.5	<2.5
22	4.40	97.40	108.40	.	7.80	11.60
23	<2.5	<2.5	<2.5	.	202.50	170.50
24	<2.5	<2.5	<2.5	.	38.30	33.00
25	<2.5	125.00	85.20	.	32.00	41.80
26	<2.5	38.50	22.20	.	<2.5	7.20
27	<2.5	95.40	81.30	.	150.40	158.00
28	<2.5	38.90	16.30	.	<2.5	<2.5
29	<2.5	40.10	38.60	.	<2.5	<2.5
29	<2.5	30.40	24.80	.	<2.5	<2.5
30	<2.5	35.40	28.80	.	178.70	198.00
31	<2.5	<2.5	6.60	.	162.20	153.10
32	<2.5	32.30	29.60	.	33.20	34.90
32	<2.5	32.80	35.30	.	32.50	32.40
33	<2.5	<2.5	<2.5	.	10.30	4.40
34	<2.5	<2.5	<2.5	.	10.30	6.10
35	<2.5	<2.5	<2.5	.	<2.5	<2.5