Toxics Testing Performance Evaluation for GB and GD

RECEIVED

APR 06 1998

OSTI

19980427 037

**Environmental Assessment Division
Argonne National Laboratory**

Operated by The University of Chicago,
under Contract W-31-109-Eng-38, for the

United States Department of Energy

DTIC QUALITY INSPECTED 3

Argonne National Laboratory

Argonne National Laboratory, with facilities in the states of Illinois and Idaho, is owned by the United States Government, and operated by the University of Chicago under the provisions of a contract with the Department of Energy.

This technical memo is a product of Argonne's Environmental Assessment Division (EAD). For information on the division's scientific and engineering activities, contact:

Director, Environmental Assessment Division
Argonne National Laboratory
Argonne, Illinois 60439-4815
Telephone (630) 252-3107

Presented in this technical memo are preliminary results of ongoing work or work that is more limited in scope and depth than that described in formal reports issued by the EAD.

Publishing support services were provided by Argonne's Information and Publishing Division (for more information, see IPD's home page: <http://www.ipd.anl.gov/>).

Disclaimer

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

Reproduced directly from the best available copy.

Available to DOE and DOE contractors from the Office of Scientific and Technical Information, P.O. Box 62, Oak Ridge, TN 37831; prices available from (423) 576-8401.

Available to the public from the National Technical Information Service, U.S. Department of Commerce, 5285 Port Royal Road, Springfield, VA 22161.

Toxics Testing Performance Evaluation for GB and GD

by H.J. O'Neill,* J.F. Schneider,* K.L. Brubaker,* T.A. Kimmell, and A.W. Anderson

Environmental Assessment Division,
Argonne National Laboratory, 9700 South Cass Avenue, Argonne, Illinois 60439

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED *ph*

MASTER

October 1997

Work sponsored by United States Department of Defense,
U.S. Army Test and Evaluation Command,
Environmental Quality Office

**O'Neill, Schneider, and Brubaker are affiliated with Argonne's Energy Systems Division,
Center for Environmental Restoration Systems.*

DTIC QUALITY INSPECTED 3

This report is printed on recycled paper.

CONTENTS

ACKNOWLEDGMENTS	viii
NOTATION	ix
ABSTRACT	1
SUMMARY	1
1 INTRODUCTION.....	3
1.1 Overview of the Toxics Testing Performance Evaluation Program.....	3
1.2 Selection of Waste Streams.....	4
1.3 Selection of Chemical Agents	6
1.4 Control Limits and Analytical Sensitivity	7
1.5 Summary of Previous Work.....	7
2 OBJECTIVES AND APPROACH	9
2.1 Overall Objectives.....	9
2.2 Chemical Considerations.....	9
2.3 Statistical Aspects of Decision-Making	10
2.3.1 False Negative and False Positive Errors	10
2.3.2 Computation of Tolerance Levels	11
2.4 Data Quality Objectives	15
2.5 General Approach.....	18
3 METHOD CERTIFICATION AND QUALITY ASSURANCE/ QUALITY CONTROL PROCEDURES	19
3.1 Method Certification Procedures	19
3.1.1 Calibration Precertification Phase	19
3.1.2 Analytical Method Certification Phase	20
3.2 Quality Assurance and Quality Control Procedures.....	21
3.2.1 Calibration Precertification Phase	21
3.2.2 Analytical Method Certification Phase	22
3.2.3 Quality Control Procedures for Sorbent Cartridges	23
3.2.4 Use of an Agent Surrogate	24
4 EXPERIMENTAL PROCEDURES	25
4.1 General Overview.....	25
4.2 Chemical Agents	25

CONTENTS (Cont.)

4.3	Preparation of Standard Solutions.....	25
4.3.1	Calibration and Check Standards	25
4.3.2	Matrix Spiking Solutions	26
4.4	Test Matrices	27
4.4.1	5% HTH Decon Fluids.....	27
4.4.2	Metal Substrate.....	29
4.4.3	Soil	30
4.5	Procedure for Spiking Sorbent Cartridges.....	30
4.6	Sorbent Cartridge Analysis by Thermal Desorption and Gas Chromatography	32
4.6.1	Thermal Desorption.....	32
4.6.2	Gas Chromatography.....	34
4.7	Confirmation Methods	34
5	RESULTS AND DISCUSSION.....	36
5.1	GB and GD Method Validation Results.....	36
5.1.1	Attainment of Data Quality Objectives	36
5.1.2	Certified Reporting Limit.....	38
5.1.3	Method Standard Deviation and Total Error.....	41
5.2	General Data Analysis.....	44
5.2.1	Treatment of Outliers	44
5.2.2	Tests for and Effects of Nonhomogeneity of Variances.....	45
5.2.3	GB and GD Percent Recovery Data	48
5.3	Summary of Quality Control Sample Results	48
6	SUMMARY AND CONCLUSIONS.....	50
7	RECOMMENDATIONS	52
8	REFERENCES.....	53
	APPENDIX A: Glossary of Terms	55
	APPENDIX B: Summary of Certification Data	57
	APPENDIX C: GB and GD Percent Recovery Investigation.....	61

TABLES

1	TTPE Waste Materials	5
2	TTPE Chemical Agents.....	6
3	Target Reporting Levels.....	7
4	Adopted TTPE Data Quality Objectives	17
5	Agent Standard Solutions in Calibration and Calibration Precertification Runs	20
6	Quality Control Samples and Criteria for Method Certification Runs.....	23
7	Calibration and Check Standard Solutions for GB and GD.....	26
8	Matrix Spiking Solutions for GB and GD.....	27
9	Typical Calibration Run Sample Sequence.....	33
10	Typical Method Certification Run Sample Sequence	34
11	Gas Chromatographic Operating Parameters	35
12	Summary of Linear Regression Results for GB and GD Found versus True Amounts.....	37
13	Summary of Method Development and Validation Results.....	38
14	Ratio of Method Standard Deviation to True Concentration: GB Data	42
15	Ratio of Method Standard Deviation to True Concentration: GD Data.....	42
16	Total Error Values: GB Data.....	43
17	Total Error Values: GD Data.....	43
18	Outlier Identification and Replacement	45
19	Tests for Nonhomogeneity of Variances.....	46
20	Summary of Method Certification Quality Control Sample Results.....	49
B.1	Certification Data for GB on Previously Noncontaminated Metal	57

TABLES (Cont.)

B.2	Certification Data for GB in Previously Noncontaminated Decon Fluid.....	57
B.3	Certification Data for GB in Suspect-Contaminated Decon Fluid.....	58
B.4	Certification Data for GB in Previously Noncontaminated Soil.....	58
B.5	Certification Data for GB in Suspect-Contaminated Soil	58
B.6	Certification Data for GD on Previously Noncontaminated Metal.....	59
B.7	Certification Data for GD in Previously Noncontaminated Decon Fluid	59
B.8	Certification Data for GD in Suspect-Contaminated Decon Fluid.....	59
B.9	Certification Data for GD in Previously Noncontaminated Soil.....	60
B.10	Certification Data for GD in Suspect-Contaminated Soil.....	60
C.1	Summary of GB Percent Recovery Data.....	66
C.2	Summary of GD Percent Recovery Data.....	66
C.3	Comparisons of GB Percent Recovery Data	67
C.4	Comparisons of GD Percent Recovery Data.....	68
C.5	GB and GD ANOVA Results	69

FIGURES

1	Frequency Distribution of Found Concentrations	12
2	Computation of Tolerance Levels	13
3	Probability of Obtaining a Found Concentration Greater Than the TL	14
4	Sorbent Cartridge Spiking Apparatus.....	31
C.1	GB Certification Data Percent Recovery versus Date.....	62
C.2	GD Certification Data Percent Recovery versus Date.....	63

FIGURES (Cont.)

C.3 GB Quality Control Data Percent Recovery versus Date.....	64
C.4 GD Quality Control Data Percent Recovery versus Date.....	65

ACKNOWLEDGMENTS

The authors would like to gratefully acknowledge the help and support of Nancy Kosko (U.S. Army Test and Evaluation Command), Carol Nudell (U.S. Army Dugway Proving Ground), and Joe King (U.S. Army Environmental Center) during the design and execution of the experimental program at Dugway.

The assistance of the following Dugway and contractor personnel in carrying out the laboratory work is also greatly appreciated: Dave Rose, Joyce Clinger, Peter Fowler, Val Harris, Jade Douglas, and Glen Wilson.

The authors would like to express their appreciation to Bill Fowler at the Southern Research Institute for his assistance in designing the program and critically reviewing the results, and to Jack Parks at Argonne National Laboratory for his help in preparing the Toxics Testing Performance and Evaluation Plan.

NOTATION

AEC	U.S. Army Environmental Center (formerly the U.S. Army Toxic and Hazardous Materials Agency, USATHAMA)
AFC	average found concentration
AL	action level
ANL	Argonne National Laboratory
ANOVA	Analysis of Variance
ATD	automated thermal desorption
CASARM	Chemical Agent Standard Analytical Reference Material
CHPPM	U.S. Army Center for Health Promotion and Preventive Medicine (formerly the U.S. Army Environmental Hygiene Agency)
CRL	certified reporting limit (method)
DEQ	State of Utah Department of Environmental Quality
DPG	Dugway Proving Ground, Utah
DQO	data quality objective
DSHW	Division of Solid and Hazardous Wastes
EPA	U.S. Environmental Protection Agency
FPD	flame photometric detector
FRPD	filtered, reduced, pH-adjusted, and diluted
GB	isopropyl methylphosphonofluoridate (Sarin)
GC	gas chromatography
GC/FPD	gas chromatography/flame photometric detector
GC/MS	gas chromatography/mass spectrometry
GD	pinacolyl methylphosphonofluoridate (Soman)
HTH	high-test hypochlorite
HVLI	high-volume liquid injection
ID	inside diameter
IRP	Installation Restoration Program
IRPQAP	Installation Restoration Program Quality Assurance Program
MDL	method detection limit
MSD	mass-selective detector
NC	never-contaminated
NC-FRPD	noncontaminated — filtered, reduced, pH-adjusted, and diluted
OD	outside diameter

PMCD	U.S. Army Program Manager for Chemical Demilitarization
ppb	parts per billion
QAP	Quality Assurance Program
QAPP	Quality Assurance Program Plan
QA/QC	quality assurance/quality control
RCRA	Resource Conservation and Recovery Act
RPD	Relative Percent Difference
rpm	rotations per minute
SC	suspect-contaminated
SC-FRPD	suspect-contaminated — filtered, reduced, pH-adjusted, and diluted
SOP	standard operating procedure.
SRI	Southern Research Institute (Birmingham, Alabama)
SWMU	solid waste management unit
TC	true concentration
TECOM	U.S. Army Test and Evaluation Command
TL	tolerance level
TMP	trimethyl phosphate (a surrogate for GB and GD)
TRL	target reporting level
TTPE	Toxics Testing Performance Evaluation (method development and validation)
UCAR	Utah Chemical Agent Rule
USATHAMA	U.S. Army Toxic and Hazardous Materials Agency (currently the U.S. Army Environmental Center, AEC)

TOXICS TESTING PERFORMANCE EVALUATION FOR GB AND GD

by

H.J. O'Neill, J.F. Schneider, K.L. Brubaker,
T.A. Kimmell, and A.W. Anderson

ABSTRACT

Residues resulting from demilitarization, treatment, cleanup, and testing of military chemical agents at Dugway Proving Ground (DPG), Utah, are currently listed as hazardous wastes by the State of Utah Department of Environmental Quality. The U.S. Army Test and Evaluation Command believes that certain categories of waste generated at DPG are not hazardous. To demonstrate this, analytical methods capable of quantitatively measuring the concentrations of chemical agents, including GB and GD, in the different waste media must be available. Argonne National Laboratory has developed methods to analyze metal substrate, spent hypochlorite decontamination fluid, and soil matrices for GB and GD. These methods involve the use of sorbent cartridge preconcentration and thermal desorption combined with gas chromatography using flame photometric detection to achieve the desired sensitivity and specificity. This report describes the methods and presents results for these three common waste matrices. The test results indicate that these methods can be used to quantitatively determine concentrations of GB and GD in the low parts-per-billion range in all sample media tested.

SUMMARY

Residues resulting from demilitarization, treatment, cleanup, and testing of military chemical agents at Dugway Proving Ground (DPG), Utah, are currently listed as hazardous wastes by the State of Utah Department of Environmental Quality, Division of Solid and Hazardous Waste. Resource Conservation and Recovery Act (RCRA) regulations allow waste generators to petition the regulator to "delist" wastes if it can be shown that they are not hazardous. The U.S. Army Test and Evaluation Command (TECOM) believes that certain categories of waste generated at DPG are not hazardous. To demonstrate this, TECOM must develop analytical methods capable of quantitatively measuring the concentrations of chemical agents, including GB (isopropyl methylphosphonofluoridate — also known as Sarin) and GD (pinacolyl methylphosphonofluoridate — also known as Soman), in the different waste media. The methods must be sufficiently sensitive to allow quantification of the agents at concentrations that pose no significant threat to human health or the environment and offer sufficient flexibility to go to lower levels if needed. TECOM obtained the assistance of Argonne National Laboratory (ANL) to develop and validate suitable analytical methods. This report describes the results of that study for agents GB and GD.

ANL has developed methods to analyze metal substrate, spent hypochlorite decontamination fluid, and soil matrices for GB and GD. These methods involve the use of sorbent cartridge preconcentration and thermal desorption combined with gas chromatography using flame photometric detection to achieve the desired sensitivity and specificity.

The methods were validated using test procedures adopted from the U.S. Army Toxic and Hazardous Materials Agency (USATHAMA) Quality Assurance Program (QAP). These test procedures demonstrate that the analytical processes can be adequately calibrated (method pre-certification phase) and that they can achieve the desired level of sensitivity and reproducibility (method certification phase). The methods described in this report meet all the requirements set forth in the USATHAMA QAP for demonstrating method validity when applied to the test media (metal substrate, spent hypochlorite decontamination fluid, and soil matrices). The test results indicate that the methods developed can be used to quantitatively determine concentrations of GB and GD in the low parts-per-billion (ppb) range in all sample media tested.

The certified reporting limits (CRLs) achieved by these methods range from 8.3 to 19 ppb for GB and from 1.8 to 5.3 ppb for GD. Method detection limits (MDLs) range from 1.7 to 8.2 ppb for GB and from 0.39 to 1.2 ppb for GD. The methods provide protection against both false positive and false negative errors in determining the level of agent with respect to an assumed action level and are acceptable with respect to a "Total Error" criterion.

The regression of found versus target agent concentrations was linear for each agent-matrix combination and all regression lines passed well within 95% confidence limits about the origin. The standard deviation of measured values about the regression line was found to increase with target concentration for all agent-matrix combinations, and in all but two cases a linear regression of the standard deviation against target spiking level was significant at the 95% confidence level or higher. The slopes of the regression lines of found versus target concentrations were statistically greater than unity at the 95% confidence level in all but one case, averaging 1.089 over all agent-matrix combinations. The reason for this apparent positive bias is unknown, although evidence suggests that it does not lie within the spiking or extraction steps.

On the basis of the findings presented in this report, we have developed the following recommendations for further work on analytical methods for GB and GD:

- The methods described in this report should be tested further on a wider variety of sample matrices, including real-world samples known to have been exposed to GB or GD and their degradation products.
- These methods should be tested at one or more independent laboratories in order to confirm their validity when implemented by different agencies.
- These methods should be modified to incorporate high-volume liquid injection (HVLI) technology that could reduce method variability and increase method sensitivity.

1 INTRODUCTION

1.1 OVERVIEW OF THE TOXICS TESTING PERFORMANCE EVALUATION PROGRAM

The State of Utah Department of Environmental Quality (DEQ), Division of Solid and Hazardous Wastes (DSHW), has designated residues resulting from the demilitarization, treatment, cleanup, and testing of military chemical agents as corrosive, reactive, toxic, and acutely hazardous wastes (Hazardous Waste No. F999). These residues are listed as hazardous waste in the State of Utah, as well as in several other states, but are not listed under the U.S. Environmental Protection Agency (EPA) regulations promulgated under the Resource Conservation and Recovery Act (RCRA). RCRA is the primary law governing management of hazardous waste in the United States.

The RCRA regulations (40 CFR 260-280), the Utah Administrative Code (R-315), and other state hazardous waste programs list specific wastes as hazardous, but they allow generators to petition the regulator to "delist," if they can demonstrate that such wastes are not hazardous. The U.S. Army Test and Evaluation Command (TECOM) believes that certain categories of F999 residues are not hazardous, and it has obtained assistance from Argonne National Laboratory (ANL) in making the delisting demonstration.

The objective of the chemical agent delisting program (hereafter referred to as the delisting program) is to demonstrate that residues resulting from materials testing activities and some remediation residues (e.g., contaminated soil) at the U.S. Army Dugway Proving Ground (DPG) in Utah should not be designated as hazardous waste. These materials have a history of exposure or possible exposure to one or more of the following chemical agents: H (HD), GB, GD, and VX. More information on the waste streams and agents, including chemical names and synonyms, is presented in Sections 1.2 and 1.3. To persuade DSHW to delist these residues, the Army must demonstrate that they do not (1) contain hazardous quantities of the listed agents; (2) contain hazardous quantities of constituents listed in 40 CFR Part 261, Appendix VIII; (3) exhibit other characteristics that would cause them to be characterized as hazardous; and (4) fail a series of acute toxicity tests. The last item is not typically required for delisting demonstrations, but it has been requested by DSHW.

The delisting program focuses on a subset of the F999 wastes generated at DPG, where the Army routinely tests the effects of military chemical agents and agent-decontamination procedures on military items. The first major objective is to develop and validate analytical methods capable of achieving the level of sensitivity required to quantitatively measure the concentrations of chemical agents in a selected set of waste matrices to be delisted. Method development and validation are not typically required for delisting because EPA has certified numerous analytical methods for use under RCRA. However, because the delisting involves substances for which EPA has not developed or certified methods, DSHW and the Army believe

it is important to show that the analytical methods to be used in the delisting process will produce data of acceptable quality.

This report presents the results of analytical method development and validation studies carried out by ANL and DPG laboratory contractor personnel at DPG as part of the Toxics Testing Performance Evaluation (TTPE) program. This program was carried out under the direction of the DPG Chemical Test Division and the TECOM Environmental Quality Division. The remainder of Section 1 provides additional introductory information. Section 2 describes the scope and objectives of this work. Section 3 summarizes the precertification and certification procedures and describes the quality assurance/quality control (QA/QC) procedures that were used. The experimental procedures used to develop and validate the analytical methods addressed in this report are described in detail in Section 4, and Section 5 presents and discusses the main results. A summary and conclusions are provided in Section 6, and Section 7 contains recommendations based on the results of this study. Section 8 lists the references used to prepare this report.

Appendix A provides a glossary of terms used in the report, and Appendix B contains summaries of the certification data for each agent/matrix combination tested. Appendix C contains results and a discussion of the agent recovery percentages obtained by using the analytical methods. The main planning document for this study, the TTPE plan, is available as a separate report (ANL 1995).

This report focuses on the TTPE program for two agents: GB and GD. ANL is preparing separate reports for HD and VX. As of the writing of this report, the remainder of the delisting program had been placed on hold. The Army is assisting DSHW in developing a new regulation for chemical agents and associated wastes, known as the Utah Chemical Agent Rule (UCAR). The UCAR is expected to affect the disposition of chemical agents and associated wastes within the State of Utah. The Army will determine whether and when the delisting program will continue as development of the UCAR proceeds.

1.2 SELECTION OF WASTE STREAMS

The DPG waste streams targeted for the TTPE program result from the performance of one of DPG's primary missions: the testing of combat equipment and chemical protective gear for functional survivability. Various equipment items, ranging from clothing and helmets to vehicles and referred to as substrates, are first exposed to one of the agents identified above. Then the items are decontaminated by using chemical decontamination fluids (decon fluids). Finally, the test items and the decon fluids are subject to post-decontamination testing and evaluation. The tested substrates that are not reused, the spent decon fluids, and related materials used in the testing are placed in containers and managed as hazardous waste.

Table 1 lists the three general categories of waste targeted for delisting. The first category consists of the tested substrates. The second category (solids) consists of soils and debris from

various locations at DPG that may have been contaminated with one or more agents during past operations. The third category (fluids) refers to spent decon fluids. One representative material from each of these three general waste categories was used in the study.

The initial focus for the substrate testing was a subset of equipment items identified as nonporous, nonpermeable solids. This subset consists primarily of unpainted metal, glassware, and ceramics. Unpainted steel was chosen to represent this subset of wastes.

The second type of waste used in this method development study is potentially contaminated soil representative of the kinds that may be identified as DPG staff remediate solid waste management units (SWMUs) through the DPG Installation Restoration Program (IRP). The specific soils used in this work are described in Section 4.6.

The third type of waste examined during this study is fluids. Personnel at DPG use various types of fluid to decontaminate test items, but one of the most often-used fluids, and hence one of the larger-volume wastes generated at DPG, is an aqueous slurry containing calcium hypochlorite [$\text{Ca}(\text{OCl})_2$]. This decon fluid is commonly prepared from a commercially available granular solid mixture containing 65% calcium hypochlorite and known as high-test hypochlorite (HTH). The most common HTH-based decon fluid used at DPG is a mixture of 15% HTH in water. This fluid can be used to decontaminate all of the agents identified in Section 1.1. By the time this fluid is used to decontaminate test items and surfaces in the test chamber and is diluted with rinse water, it contains no more than 5% HTH. A 5% HTH decon fluid was identified as the most suitable for use in analytical method development because it is most representative of the waste decon fluid generated at DPG.

TABLE 1 TTPE Waste Materials

Category	Type of Material
Substrates	Unpainted metal
	Alkyd-painted metal
	Polyurethane-painted metal
	Butyl rubber
	Neoprene rubber
	Chemically protective fabric
	Polystyrene
	Polymethyl methacrylate (Plexiglas)
Solids	Soil, debris
Fluids	Previously used decon fluid

Two types of soil and decon fluid sample matrices were used to demonstrate the performance of the analytical methods. Never-contaminated (NC) soil samples consisted of "synthetic" waste samples, prepared by blending soils collected at DPG from areas known never to have been exposed to agents. Suspect-contaminated (SC) soil samples consisted of samples of soil collected from an area at DPG believed to have been previously exposed to chemical agents. SC decon fluids were prepared from HTH decontamination solutions in accordance with DPG protocols by adding agent (GB or GD) to fresh NC (never used) HTH decon solutions.

NC soil and decon fluid samples were used in addition to SC samples because SC samples may contain agent degradation products that could interfere with the measurement of agent concentrations, so the use of both NC and SC samples in the method verification effort is considered prudent and essential.

For the substrate testing, only NC metal samples were used because the decontamination of unpainted metal was believed to be extremely effective so the possibility that significant differences might exist between NC and SC unpainted metal samples was considered negligible.

1.3 SELECTION OF CHEMICAL AGENTS

The wastes of interest at DPG may have been exposed to one of several chemical agents. Table 2 lists the agents of concern for the TTPE program. GB, GD, and VX act primarily on the nervous system by inhibiting the enzyme acetylcholinesterase and are referred to as nerve agents. H/HD, a vesicant or blister agent, is a cytotoxin that attacks virtually all flesh, although the eyes and the mucous membranes in the respiratory tract are particularly vulnerable.

Decontamination of tested substrates rapidly converts agents into relatively nontoxic components. For example, the G agents (e.g., GB, GD) are rapidly hydrolyzed by mixing with strong caustics, such as sodium hydroxide. Decon fluids containing hypochlorite, such as HTH, rapidly oxidize HD as well as the G agents, VX, and other organic compounds. The chemistry of

TABLE 2 TTPE Chemical Agents

Acronym	Chemical Name	Synonym	CAS Registry Number
GB	Isopropyl methylphosphonofluoridate	Sarin	107-44-8
GD	Pinacolyl methylphosphonofluoridate	Soman	96-64-0
H	Bis-(2-chloroethyl) sulfide, 70% HD by weight	Mustard, technical grade	N/A
HD	Bis-(2-chloroethyl) sulfide	Mustard, distilled	505-60-2
VX	O-ethyl S-(2-diisopropylaminoethyl) methylphosphonothiolate	None	50782-69-9

the decontamination reactions of each agent is discussed in depth by Yang et al. (1992) and by Rosenblatt et al. (1996). In addition, much of the possible contamination of soil by agents in the past (designated F999) was caused by disposal of decontaminated substrates rather than disposal of actual agent materials. In most cases, these residues and soils have been exposed to the elements for many years; any agent that was initially present is expected to have degraded into relatively nontoxic components long ago. These waste materials are therefore ideal candidates for delisting.

1.4 CONTROL LIMITS AND ANALYTICAL SENSITIVITY

As stated above, the analytical methods to be used in delisting must show that the wastes of interest do not contain hazardous concentrations of the listed agents. Action levels (ALs), or the concentration below which a given agent does not pose a significant hazard, are being established for each agent in the wastes to be delisted. The analytical methods used for agent analyses must be sensitive enough to provide quantitative results at agent concentrations below the ALs.

ANL worked under the direction of the U.S. Army Center for Health Promotion and Preventive Medicine (CHPPM, formerly the U.S. Army Environmental Hygiene Agency) to establish the ALs. As discussed in Section 1.1, this work was placed on hold. Approved ALs have therefore not yet been established. In the interim, target reporting levels (TRLs) for the analytical methods have been set at the levels shown in Table 3.

1.5 SUMMARY OF PREVIOUS WORK

The initial phase of the delisting program was carried out by Southern Research Institute (SRI) from 1991 to 1993. During this period, SRI gathered, evaluated, and documented pertinent information concerning agents and waste streams and other supporting data. SRI also prepared a series of background documents and demonstration test plans for use in the delisting program.

Several of SRI's documents are relevant to the work described in this report. SRI's *3X/5X Delisting Plan — Analytical Methods Report* (SRI 1992) contains a review of the analytical methods for chemical agents that were considered most likely to be useful for the TTPE program. The methods described in the report, presently in use at two different U.S. Army installations, fall into two categories: soil and substrate analysis and fluid analysis.

TABLE 3 Target Reporting Levels

Waste Category	GB	GD
Fluids ($\mu\text{g/mL}$)	0.024	0.006
Soil ($\mu\text{g/g}$)	0.024	0.006
Substrates ($\mu\text{g/g}$)	0.024	0.006

The first category comprises methods recommended by SRI for the analysis of chemical agents in soil and substrates to support delisting. The methods were developed by DPG personnel for the analysis of agents in soil and are presently used by DPG in its IRP. Analytical methods for groundwater were also developed by DPG and adopted for the IRP. SRI suggested that the DPG methods be modified by adopting a sorbent cartridge preconcentration and thermal desorption methodology if greater sensitivity is required.

The second category consists of methods recommended by SRI for the analysis of matrices derived from decon fluids. These methods were developed by SRI for the Monitoring Branch of the Army's Program Manager for Chemical Demilitarization (PMCD) and were intended for use in analyzing scrubber effluent solutions and bleach, fluids similar to DPG's decon fluids. In the PMCD methods, these fluids are first treated with a reducing agent to destroy hypochlorite (which would be present in HTH solutions) and then pH-adjusted before extracting and analyzing for agent. The extraction steps in the PMCD methods are identical to those in use at DPG for analyzing agents in groundwater, although PMCD used the sorbent cartridge preconcentration and thermal desorption methodology in its analyses.

Another relevant SRI document is the *Waste Sampling Test Report*, which describes a series of spiking and recovery experiments conducted by SRI. In these studies, all of the substrates listed in Table 1, Section 1.3 of this report, except for unpainted metal, were spiked with the agents GB, GD, and HD at levels in the low parts-per-billion (ppb) range. The spiked substrates were then extracted and analyzed for agent.

In SRI's studies, the methods used for extraction and analysis were quite similar to those used for soil analysis at DPG, but they employed a sorbent cartridge preconcentration and thermal desorption methodology as part of the analytical process. Although SRI's methods were not optimized for maximum sensitivity, this methodology significantly increased analytical sensitivity and reduced problems caused by analytical interferences. In most of the cases presented, SRI showed that these agents could be reproducibly extracted from these substrates and analyzed to achieve low detection limits. Further, SRI's results show that incorporation of the sorbent cartridge preconcentration technique into the analytical process provides excellent analytical recoveries for these agent/substrate combinations.

SRI's *Analytical Methods Test Plan for the 3X/5X Delisting Plan* (SRI undated) is also relevant to the method validation effort described here, in that portions of the TTPE plan were based on the preliminary work reported by SRI.

2 OBJECTIVES AND APPROACH

2.1 OVERALL OBJECTIVES

As indicated in Section 1, the objective of the work described here was the development and validation of analytical methods for GB and GD suitable for use in measuring low levels of these agents in several DPG waste matrices. These methods must be capable of achieving acceptable levels of specificity and sensitivity for selected classes of wastes, and their suitability must be demonstrated by an appropriate validation process.

The terms "method validation" and "method certification" are often used interchangeably. According to Taylor (1987), "Validation is the process of determining the suitability of a measurement system for providing useful analytical data." Accordingly, a method may be considered validated when sufficient evidence is presented that it is suitable for its intended applications. It is impossible to anticipate all potential technical problems that might arise when a particular method is used for analysis of real-world samples; in this sense, method validation is necessarily always incomplete.

Method certification involves demonstrating that an analytical method meets certain prescribed criteria set forth by the certifying agency. The testing procedures adopted in this study are based on those contained in the U.S. Army Toxic and Hazardous Materials Agency (USATHAMA) Quality Assurance Program Plan (QAPP) (USATHAMA 1990) for its method certification program. By meeting the USATHAMA criteria for method certification, ANL and DPG personnel provided sufficient evidence to demonstrate that the methods developed can be applied to actual metal, decon fluid, and soil samples to be tested in the future.

2.2 CHEMICAL CONSIDERATIONS

Validating the methods used to determine the amount of chemical agent in the soil and decontaminated substrate matrices is relatively straightforward. However, validating the methods to detect chemical agents in the spent decon fluid matrix is problematic because the decon fluid is designed to destroy chemical agent on contact. Standard chemical analyses of spiked decon fluid would not be expected to detect the agents because they would have been quickly destroyed. However, it is important to demonstrate that the analytical methods developed during this effort would detect any agent that, for whatever reason, may be present in the spent decon fluid. EPA ran into a similar problem during a delisting action involving cyanide. The waste matrix of concern contained a strong oxidizing agent (i.e., bromine) that destroyed any cyanide that may have been present in the waste or that was spiked into solution for method testing purposes. In this case, EPA determined that the waste could be delisted if the petitioner was able to demonstrate, among other things, that the cyanide spiked into a waste sample could be

quantitatively recovered from the waste if the oxidizing agent was first removed from the waste matrix (through addition of a reducing agent) (Kayser 1988).

A similar approach was adopted in this study. In the preparation of decon fluid samples, sufficient ascorbic acid (vitamin C) is added to each batch of HTH decon fluid to completely react with and destroy the active oxidizing agent in the fluid (hypochlorite ion). During the spiking and extraction operations conducted as part of the sample analysis, agent is added to the neutralized fluid; the agent can then be extracted and analyzed without being destroyed. The specific procedures used are discussed in Section 4.4. The results obtained demonstrate the validity of the analytical method for the decon fluid matrix.

2.3 STATISTICAL ASPECTS OF DECISION-MAKING

Analytical results obtained using the methods developed in this study are intended to be used to determine whether certain wastes pose a sufficient hazard to warrant their regulation (or deregulation) as hazardous waste. Some statistical aspects of the decision-making process are outlined in this section to put the results of this method development and validation study into perspective and to indicate the context in which measures of method performance should be considered.

2.3.1 False Negative and False Positive Errors

The decision regarding whether to delist a particular waste was intended to be based on a comparison of an estimated agent concentration in the waste with the appropriate AL. If the estimated concentration is greater than or equal to the AL, the waste should not be delisted; if the concentration is less than the AL, the waste can be delisted. The estimated agent concentration will be obtained using one of the analytical methods developed in this study.

The decision regarding whether to delist the waste could be wrong for one of two reasons. If the agent concentration is higher than the AL, but because of unavoidable random errors the chemical analysis gives a concentration below the AL, the waste would be incorrectly delisted; this is called a "false negative." Conversely, if the true agent concentration is less than the AL, but the analysis gives a concentration greater than the AL, the agent would not be delisted even though it could be. This is called a "false positive."

The possible occurrence of false negatives is of significant concern because wastes not satisfying the safe concentration criterion could be incorrectly delisted, posing a potential threat to human health or the environment. In making decisions of this type on the basis of a comparison of analytical results with an action level, the possibility always exists that a false negative error could be made. The decision-making process must be structured so that the probability of such an error does not exceed an acceptable limit. An appropriate and useful way

to accomplish this is to compare the analytical result not with the AL but with a lower value, the Tolerance Level (TL), the value of which is chosen to provide the necessary safeguard against false negative decisions. It is suggested that the TL be defined so that the maximum probability of making a false negative decision when the true concentration is equal to the AL is 2.5%; for higher concentrations, the false negative probability would be very much lower — negligible for agent concentrations significantly higher than the AL. So, using this approach to decision making the analytical result should be compared with the TL, and the waste delisted if the concentration is less than the TL. The computation of the TL in terms of method performance parameters is discussed in Section 2.3.2.

Minimizing the probability of a false negative error is considered more important than minimizing the probability of a false positive because of the potential for adverse human health or environmental consequences arising from a false negative error. Defining and using a TL in the manner described ensures that the maximum probability of a false negative error equals the suggested value of 2.5% (0.025).

The possible occurrence of false positives is also of concern, though, because wastes that could be delisted and managed using much less expensive but still environmentally protective methods must be treated as hazardous waste. The goal of minimizing the probability of false positives is somewhat incompatible with that of minimizing false negatives. In order to satisfy the suggested 2.5% probability criterion for the occurrence of false negatives when the agent concentration in the waste equals the AL, it is necessary to allow a relatively high probability of false positives for true concentrations just below the AL. For lower concentrations, the probability of false positive results decreases sharply. The range in which the potential for false positive results is high can be made smaller by decreasing the method standard deviation, either by method improvements or by analyzing replicate samples and averaging to obtain a more precise concentration estimate. To a limited extent, basing the estimated standard deviation on a larger number of measurements would also help. The actual incidence of false positives will depend on the actual agent concentrations encountered. If, as expected with the subject waste matrices, most agent concentrations are below analytical detection limits, the incidence of false positives will be very small.

2.3.2 Computation of Tolerance Levels

Figure 1 shows a theoretical plot of the probability distribution for agent concentration values estimated by using an analytical method ("found" concentrations). The probability distribution gives the relative frequency of occurrence of all possible found concentrations. Because of possible systematic errors, the overall average found concentration may differ from the true concentration; individual estimates will differ from the average found concentration because of unavoidable random errors. An individual result will therefore differ from the true concentration because of both random and systematic errors. The difference between the overall average found concentration (AFC) and the true concentration (TC) is called the method bias, or

FIGURE 1 Frequency Distribution of Found Concentrations

just the bias (B); note that the bias can be either positive or negative. In Figure 1, B is negative, corresponding to the illustrated negative bias. B is defined by the following: $B = AFC - TC$. The occurrence of random errors of various magnitudes is characterized by the method standard deviation σ , which is a measure of the width of the distribution shown in Figure 1. For any given method, only estimates of the bias and standard deviation are available.

The computation of the TL is illustrated in Figure 2. The TC equals the action level, and the TL is set at the value at which the area under the distribution curve and to the left of the TL line equals the desired probability 0.025. This value can be computed by using Equation (1):

$$TL = AL + B - tS \quad (1)$$

FIGURE 2 Computation of Tolerance Levels

In this equation, S denotes an estimate of σ and t denotes the appropriate numerical factor associated with the desired false negative probability. If σ is known, then S equals σ and the appropriate value of t is 1.960. If only an estimated S value based on n measurements is available, the appropriate value of t is the value for which the cumulative Student's- t distribution with $n-1$ degrees of freedom equals 0.975 (Snedecor and Cochran 1989). If the estimate of S is based on four measurements, t equals 3.182. The appropriate value of t is also equal to the critical two-tailed 95% confidence value for a Student's- t variable with $n-1$ degrees of freedom, a quantity that is more frequently tabulated than the cumulative Student's- t distribution.

Figure 3 shows the probability that the found concentration would be greater than the TL, i.e., that the corresponding waste sample would be rejected as a candidate for delisting, as a function of the true agent concentration TC (more precisely, as a function of the ratio TC/AL).

FIGURE 3 Probability of Obtaining a Found Concentration Greater Than the TL

This probability also depends on the estimated method standard deviation, and curves are shown for three selected values of S . To the left of the $TC/AL = 1$ line, the curves show the probability of a false positive decision because they show the probability of a found concentration being greater than the TL when TC/AL is less than one. In generating these curves, we have assumed that the estimate S is based on four measurements. If more than four measurements are used in obtaining the estimate S , the curves would be steeper and closer to the $TC/AL = 1$ line. These curves must all pass through the value 0.975 when the TC equals the delisting level so that the maximum probability of a false negative is 0.025.

Qualitatively, the curves show that as the TC gets smaller, the chances of making a false positive error also get smaller, as one would expect. They also show that if the TC is close to the AL , the chances of making a false positive error increase; this is also what would be expected, because the analytical method cannot unquestionably distinguish the TC from the AL in such a

case. The figure also shows that there will always be a region just below the AL in which the probability of a false positive cannot be reduced in practice because the estimated method standard deviation S cannot be made arbitrarily small. For example, assume that the target maximum probability for false positive errors is also 0.025. Then, if S equals 0.025 times the AL, the corresponding curve on Figure 3 shows that for TCs between $0.85 \times \text{AL}$ and AL, the probability of a false positive exceeds the target value of 0.025, but for concentrations less than this, the chances of a false positive are acceptable.

As indicated above, the curves shown in Figure 3 can be improved to a limited extent simply by basing the estimate S on more than four measurements, reducing the size of the region in which the probability for false positive results is high. The improvement would result from the fact that a higher level of confidence can be placed in an estimate S that is based on more measurements. As the number of measurements increases, the estimate S will gradually approach the value σ . The gain that can be achieved this way is limited, however, by the value of σ . For example, if σ is equal to $0.025 \times \text{AL}$ and if the estimate S is based on a sufficiently large number of measurements, the region for which the false positive probability exceeds 0.025 would extend from approximately $0.9 \times \text{AL}$, instead of $0.85 \times \text{AL}$, to AL. The only way to further minimize the range in which the chances of a false positive are greater than 0.025 is to reduce the magnitude of S (either by method improvements or by taking replicate samples and averaging). These considerations, together with the definition of the TL, show the importance of minimizing the method standard deviation.

2.4 DATA QUALITY OBJECTIVES

This section presents and discusses the data quality objectives (DQOs) adopted for the analytical method development/validation effort described in this report. The DQOs address several method performance requirements.

Three general performance criteria for analytical methods address sensitivity, linearity, and blank correction. The methods developed for this program must be sufficiently sensitive that agent concentrations below the AL can be detected and measured quantitatively. One measure of method sensitivity that has been used by the Army is the "Certified Reporting Limit" or CRL (USATHAMA 1990). Based on the description provided in Section 14 of the USATHAMA QAPP, the CRL is the true concentration at which there is less than a 5% chance of obtaining a found concentration below the value that can just be distinguished from zero at the 95% confidence level. The CRL was computed using the Army's Installation Restoration Program Quality Assurance Programs (IRPQAP) software (Potomac Research, Inc. 1990) or equivalent. The corresponding DQO for each method developed in this program was that the CRL be below the associated Target Reporting Limit. The Method Detection Limit used by EPA was also computed and reported for each method. This measure of sensitivity is defined as "the minimum concentration of a substance that can be measured and reported with 99% confidence that the analyte concentration is greater than zero and is determined from analysis of a sample in a given matrix type containing the analyte" (EPA 1986).

In addition to being sufficiently sensitive, analytical methods for each matrix should give results that are linearly dependent on the true agent concentration without an unacceptable bias at zero concentration. Tests for linearity and for the consistency of results with a zero intercept were also performed by using the IRPQAP software or equivalent. The test for linearity evaluates the results for a lack of fit with a linear regression model and is sensitive to systematic deviations from such a model. The zero-intercept test is equivalent to computing the 95% confidence limits on the regression intercept and checking to determine whether they enclose the value zero. The associated DQOs were that each method must pass the lack-of-fit and zero-intercept tests. In addition, in what the USATHAMA QAPP terms the "precertification phase," the researchers must show that the assumed calibration model for each method is appropriate, and the method must provide results without significant zero correction. These checks were made using calibration samples, in which the analytes are spiked into a suitable solvent rather than into the target sample matrices. Again, the tests were performed using the IRPQAP software or equivalent, and the associated DQOs were that each method successfully pass the tests.

As discussed in Section 5.1.1, the IRPQAP software is not designed to perform the lack-of-fit and zero-intercept tests when one or more outliers have been deleted or replaced in the method certification dataset. These tests were performed and the CRL was calculated independently in all cases to check the accuracy of the IRPQAP results and to provide the correct results for those cases in which outliers were identified.

As described in Section 2.3, the magnitude of the method standard deviation (S) relative to the delisting level (AL) affects the probability of a false positive error. Although the true incidence of false positives will depend strongly on the actual agent concentrations encountered, a method for which S/AL is below a specified maximum will help ensure that the incidence of false positives is acceptably small. The adopted DQO is that, for each method, S/AL should be less than 0.1 for spiking levels at or near the TRL, including spiking levels at 0.5, 1.0, and 2.0 times the TRL. Specifying these spiking levels from among those that are used allows for the possibility that the standard deviation may increase with spiking level; only the method standard deviation near and below the AL or the TRL is significant for minimizing the possibility of false positives.

The scientific community has proposed and used a variety of ways to measure the performance of trace analyses methods. One such measure is "total error," proposed by McFarren, et al. (1970). This parameter was originally defined on a heuristic basis and was intended to provide a rational and non-arbitrary measure to compare a very wide spectrum of analytical methods. The definition of "total error" incorporates both accuracy, as measured by the method bias, and precision, as measured by the method standard deviation, into a single numerical measure of performance. Three levels of method acceptability — excellent (total errors less than 25%), acceptable (total errors between 25% and 50%), and unacceptable (total

errors greater than 50%) — were defined. McFarren et al. originally proposed that the total error be computed using the following equation:

$$\text{Total error} = \frac{\text{Absolute value of mean error} + 2 (\text{Std. dev.})}{\text{True value}} \times 100\% \quad (2)$$

However, the factor 2 that appears in this equation approximately represents the two-tailed 95% confidence critical value for an assumed normal distribution for the measurement errors. If the method standard deviation (S) is based on only a limited number of measurements, the factor 2 should, in principle, be replaced by the appropriate Student's-t critical value. In this study, at each spiking level, S is based on four measurements, and the associated 95% confidence critical value is 3.182. The total error is computed in this study using the above equation with the factor 2 replaced by the value 3.182. The corresponding DQO is that, at each spiking level, the total error for each method should be less than 50%. In practice, because the use of the factor 2 is widespread in the scientific community even when the total error is based on a limited number of measurements, values for the total error using both factors are computed and reported for each method in this study.

The DQOs adopted for each method are summarized in Table 4. The first three DQOs listed are absolute requirements; if a method fails to satisfy them, the reasons for the failure are to be determined and corrected and the certification analyses re-run. The last two DQOs are not absolute requirements, but are desirable targets. Method modifications can be made based on the degree to which these targets are met (or not met).

EPA has recently issued guidance for developing and validating methods intended to be candidates for inclusion in EPA's compendium of analytical methods for determining compliance under RCRA (EPA 1995). Although this guidance was not used directly in

TABLE 4 Adopted TTPE Data Quality Objectives

1	The method CRL must be less than the TRL.
2	The method must pass lack-of-fit and zero-intercept tests for calibration (precertification) samples.
3	The method must pass lack-of-fit and zero-intercept tests for target sample matrices (certification samples).
4	The ratio of the method standard deviation to the TRL should be less than 0.1 for spiking levels at or near the TRL.
5	The "total error" for each spiking level should be less than 50%.

developing the DQOs presented in this section, it is useful to present EPA's requirements for formal method validation for comparison and future reference. Regarding method sensitivity, EPA merely states that a method must exhibit analytical sensitivity appropriate for its intended application and does not specify numerical criteria. EPA guidance establishes the acceptable spike recovery range at 80% to 120% for method accuracy and precision, and a maximum relative percent difference between spike duplicates of 20%. EPA also addresses the issue of repeatability, which is measured as long-term precision.

2.5 GENERAL APPROACH

In order to achieve the necessary specificity and sensitivity, ANL employed gas chromatographic methods in combination with analyte preconcentration techniques and the most sensitive available detectors. Preliminary estimates of the instrumental detection limits for flame photometric detectors operated in the phosphorus mode indicated that the use of gas chromatography (GC) with flame photometric detection (FPD) would provide more than adequate sensitivity for GB and GD. These expectations were borne out by the experimental results obtained in this study.

In order to ensure that the necessary sensitivity would be achievable for GB and GD, sorbent cartridge preconcentration and thermal desorption were used to introduce the sample into the gas chromatograph. This approach involves transferring a relatively large portion of sample extract into a sorbent cartridge in which the agent is trapped on a suitable sorbent resin. Care is taken not to inject liquid extract directly into the sorbent cartridge because the solvent could damage the resin as well as wash the agent off the resin and through the tube. The technique used to introduce the sample extract into the sorbent cartridge involves the use of a second tube, called a "fuzz tube," that contains silanized glass wool and is directly connected to the sorbent cartridge. The solvent extract is dosed onto the glass wool in the fuzz tube, and a stream of air is pulled through the two tubes. The solvent and agent are volatilized by the air flowing through the glass wool and are swept into the sorbent cartridge in the vapor phase. The solvent passes through the sorbent cartridge, but the agent is trapped on the resin.

The volume of extract that can be processed in this way is one to two orders of magnitude larger than the volume that could be directly injected into the GC, with a corresponding increase in sensitivity. The agent and any other volatile substance trapped on the sorbent cartridge are then thermally desorbed onto the GC column, separated chromatographically, and detected. This technique provides a way to adjust the method sensitivity over a wide range down to very low levels. In addition, any non-volatile sample or matrix constituents that are extracted along with the agent are permanently deposited in the fuzz tube. They are not introduced into the GC and cannot interfere with its operation.

Section 4 provides a complete description of the equipment, instrumentation, and experimental procedures used in this study.

3 METHOD CERTIFICATION AND QUALITY ASSURANCE/ QUALITY CONTROL PROCEDURES

In this section, the procedures required by the USATHAMA QAPP for certification of analytical methods are described along with the quality control (QC) procedures that ANL employed during this study.

3.1 METHOD CERTIFICATION PROCEDURES

The methodology described in the USATHAMA QAPP formed the basis for the method validation process used in this study, so that the adopted procedures were compatible and consistent with those specified in that document. The QAPP organizes method validation into two phases: calibration precertification and method certification. The calibration precertification phase addresses the ability of the analytical measurement system to provide reproducible results proportional to the true analyte concentrations in standard solutions. The method certification phase addresses the performance of the method on target spiked matrices with respect to sensitivity, linearity, and bias. These two phases are discussed in Sections 3.1.1 and 3.1.2, respectively.

3.1.1 Calibration Precertification Phase

The analytical methods developed in this study require the use of independently generated calibration curves. In such a case, the assumed calibration model must be verified prior to conducting the actual method certification runs. To do this, a suitable set of experiments, as specified in the USATHAMA QAPP, must be carried out. It was expected that the calibration curves for GB and GD would be linear with no significant bias at zero concentration. To confirm these expectations, the QAPP required the examination of a minimum of two sets of calibration data generated on successive or nearly successive days. The calibration samples run during the precertification phase consisted of a single agent spiked into solvent at concentrations determined by the assumed TRL.

The TRL is a concentration value that serves as a preliminary lower bound to the range of concentrations for which the method is expected to yield quantitative results. The TRL guides the selection of appropriate spiking levels in both the precertification and method certification phases. The TRL values adopted for use in this study are given in Table 3. The USATHAMA QAPP requires that precertification data be obtained over a wider range of concentrations than will be used for method certification: the lowest concentration used must be 25% less than $0.5 \times \text{TRL}$, and the highest must be 25% higher than $10 \times \text{TRL}$. The concentrations chosen for use in the precertification phase corresponded to the following concentrations in each matrix to be studied: $0.375 \times \text{TRL}$, $0.5 \times \text{TRL}$, $1 \times \text{TRL}$, $2 \times \text{TRL}$, $5 \times \text{TRL}$, $10 \times \text{TRL}$, and $12.5 \times \text{TRL}$.

Table 5 shows the agent concentrations and the amount of agent ideally transferred onto the GC column for each calibration solution. For a given agent, the target concentration range was essentially the same for all matrices, and the sample workup procedures for each method for a given agent were designed so that the same calibration curve could be used for all analyses for the specified agent in different matrices. For example, analyses of GB in soil, on unpainted metal, and in decon fluid at concentrations equal to the corresponding TRL would all theoretically deliver the same amount of GB (4.8 ng) to the gas chromatograph. Because of this, only one calibration precertification step was needed for each agent, rather than one for each agent/matrix combination. For each agent, four sets of precertification calibration data instead of two were analyzed using the U.S. Army IRPQAP software, which checked the data for lack of fit with the assumed linear model and for the absence of a statistically significant non-zero intercept at zero agent concentration. IRPQAP also required that the value of R^2 for each individual calibration curve be no lower than 0.995. (R^2 is a statistical measure of the quality of the fit between the measurement results and the true values, based on an assumed linear relationship.) Modifications to the basic analytical procedures used for each agent were made as needed until a successful IRPQAP analysis indicated that precertification had been successfully completed.

3.1.2 Analytical Method Certification Phase

For the method certification phase, the QAP requires that analyses be conducted on sets of spiked sample matrices on four successive or nearly successive days. An initial

TABLE 5 Agent Standard Solutions in Calibration and Calibration Precertification Runs

Calibration Standard Solution	Agent and TRL (ng/g or ng/mL) ^a			
	GB (24)		GD (6)	
	Conc. ^b	Amount ^c	Conc. ^b	Amount ^c
0.375×TRL	0.18	1.8	0.045	0.45
0.5×TRL	0.24	2.4	0.060	0.6
1.0×TRL	0.48	4.8	0.12	1.2
2.0×TRL	0.96	9.6	0.24	2.4
5.0×TRL	2.4	24.0	0.60	6.0
10.0×TRL	4.8	48.0	1.2	12.0
12.5×TRL	6.0	60.0	1.5	15.0

^a Units for the TRL are ng/g for soil and substrate matrices and ng/mL for decon fluids.

^b Units for the solution concentrations are ng/μL.

^c Amount of agent transferred to the GC in a 10-μL injection, in ng.

calibration run was carried out on the first day of each week and on additional workdays as needed. The calibration curve was considered acceptable if the associated value of R^2 was equal to or greater than 0.995 and if the intercept was statistically indistinguishable from zero.

The spiking levels used in the method certification runs are specified by the QAP in terms of the TRL: $0.0 \times \text{TRL}$, $0.5 \times \text{TRL}$, $1 \times \text{TRL}$, $2 \times \text{TRL}$, $5 \times \text{TRL}$, and $10 \times \text{TRL}$. The data from these analyses for each method were analyzed, in part, using the IRPQAP software, which checked the data for lack of fit with the assumed linear model and for the absence of a significant bias at low concentration and computed the CRL.

The QAP also requires that the data for each method be checked for outliers, which are individual results that do not conform to the rest of the data as indicated by a statistical test. This test is not carried out by the IRPQAP software. If too many outliers are found, one or more of the certification runs must be redone. It is important to screen the data for outliers, because they can significantly affect the results of the method certification computations. Modifications to the general USATHAMA procedure, as described in Section 5.2, were made for this effort.

Method certification was considered successful if, after we checked the results for outliers and replaced any that were found with appropriately computed values, the dataset passed the tests for linearity and zero intercept and the computed CRL was less than the specified TRL.

3.2 QUALITY ASSURANCE AND QUALITY CONTROL PROCEDURES

3.2.1 Calibration Precertification Phase

In the calibration precertification process, several QC samples were included with each set of calibration samples. A check standard, prepared from a different stock solution by an analyst other than the one preparing the rest of the solutions, was included. The check standard concentration in each case corresponded to $8 \times \text{TRL}$ in the sample matrices to be studied. In addition to the check standard, two types of blanks were routinely run during calibration precertification runs. The first type utilized pure solvent only; typically, two such solvent blanks were included in each precertification run — one at the beginning of the run and the other after all spiked samples were run. The second type involved drawing only air through the fuzz tube and the thermal desorption tube, then analyzing the thermal desorption tube in the usual way. One such air blank was analyzed for each calibration or calibration precertification run. The results of these QC analyses are discussed in Section 5.3.

3.2.2 Analytical Method Certification Phase

Several types of QC samples, listed in Table 6, were included in each method certification run. These samples were designed to check for the presence of interfering substances or cross contamination, to check for acceptable method reproducibility, to ensure that the method calibration was accurate and reproducible, and to detect undesirable or irreproducible matrix extraction effects.

Matrix blank samples were samples of the matrix under investigation that had not been spiked with agent. The gas chromatograms from these analyses were examined visually for peaks eluting at the same retention time as the agent.

One matrix spike sample at the 10×TRL level was included automatically as part of each certification run. A second matrix spike sample at the same spiking level was included in the run; a comparison of the two results served as a check on method reproducibility at the high end of the concentration range.

Two solvent spike samples at the 1×TRL level were run in order to directly assess reproducibility at that specific agent concentration and to identify differences with matrix samples spiked at the same level. A solvent spike sample consisted of a suitable volume of the extraction solvent, chloroform, injected with the appropriate amount of agent.

Finally, two calibration standards — the low standard (0.375×TRL) and the high standard (12.5×TRL) — and the check standard (8×TRL) were included in the certification run to verify the reproducibility and accuracy of the calibration curve used.

In Table 6, RPD denotes the relative percent difference between two values, computed according to the following formula:

$$RPD = \frac{|Value\ 1 - Value\ 2|}{\left(\frac{Value\ 1 + Value\ 2}{2}\right)} \times 100\% \quad (3)$$

When four method certification runs for a given agent/matrix combination were completed, the resulting dataset was checked for the presence of outliers. The procedures used for this check are discussed in Section 5.2.1. The QAPP requires that no more than one value at a given spiking level and no more than a total of two values be rejected as outliers from a dataset. If these criteria are not met, the offending runs are to be repeated. In addition, if two outliers occur on the same day, that day's certification run must be repeated.

TABLE 6 Quality Control Samples and Criteria for Method Certification Runs

QC Sample Name	Spiking Level	Acceptance Criteria		
Blanks and Duplicates				
Matrix Blank (two used)	0×TRL	No discernible peaks		
Matrix Spike Duplicates	10×TRL	RPD <= 15%		
Solvent Spike Duplicates	1×TRL	RPD <= 15%		
Calibration and Check Standards		Calibration Run		Earlier Run
		Same Day	Earlier Day	Same Day
Low Standard	0.375×TRL	RPD<= 15%	RPD<= 25%	RPD<= 15%
High Standard	12.5×TRL	RPD<= 15%	RPD<= 25%	RPD<= 15%
Check Standard	8×TRL	RPD<= 15%	RPD<= 25%	RPD<= 15%

After checking the dataset for outliers and replacing any outliers found with values computed as described in Section 5.2.1, the data were analyzed using IRPQAP and other software. A visual check was made to prevent transcription errors during data entry into IRPQAP or into spreadsheets designed and used for additional statistical analyses.

3.2.3 Quality Control Procedures for Sorbent Cartridges

All sorbent cartridges obtained from the supplier came marked with a unique serial number. When received, each cartridge was visually examined to verify the integrity of the packing and to detect any chipped or broken ends. If a cartridge passed the visual examination, it was entered in the Sorbent Cartridge Log, which contained the following information: (1) the date received; (2) serial number; (3) type of resin, mesh size, and supplier; (4) visual inspection results; (5) pressure drop test results (see below); (6) name of the individual performing the pressure drop test; and (7) date of the pressure drop test. A separate individual cartridge usage logsheet was used to track the usage of the cartridge throughout its lifecycle. The usage logsheet contained the date (each time the cartridge was used), the agent for which the cartridge was used, and the date of each reconditioning. A separate set of cartridges was used for each agent. A physical description of a sorbent cartridge is provided in Section 4.5.

Once a cartridge had been logged in, the pressure drop across the cartridge was measured while drawing ambient air through it at a flow rate of 500 mL/min. The pressure drop across a similar tube containing glass frit, glass wool, and the spring clip but no sorbent resin was also measured and subtracted from the pressure drop for the complete cartridge. The cartridge was accepted if the corrected pressure drop was in the range of 182 ± 5 mmHg.

During routine operations, the sorbent cartridges were reconditioned automatically by the process of thermal desorption conducted as part of an analysis. However, prior to reuse, each cartridge was reconditioned again using a separate ATD unit at a temperature of 275°C and a flowrate of 80 mL/min for 1 h. In addition, 10% of the reconditioned cartridges were randomly selected and run as blanks to check for any residual material that might interfere with the analyses.

If at any time a sorbent cartridge showed signs of discoloration, resin separation, resin degradation, breakage or chipping, leakage in the ATD, artifacts in the gas chromatograms, plugging, or loss of resin, the cartridge was removed from service, the resin was removed, and the cartridge cleaned and repacked. After repacking, the cartridge was subject to the qualifying procedure described above.

3.2.4 Use of an Agent Surrogate

During the spiking and extraction experiments, each sample was spiked with trimethylphosphate (TMP) in addition to the agent under investigation (GB or GD). TMP was used as a surrogate, a substance which is chemically stable but is similar in structure and behavior to both GB and GD and can be detected by FPD. TMP was always spiked at the 1×TRL level regardless of the agent concentration. The surrogate peak in the gas chromatogram obtained from the analysis was examined visually and compared qualitatively with previous results. The purpose of the surrogate was to provide a QC check for unusual matrix effects or gross errors in the spiking and solvent extraction steps of the analytical process. No calibration was carried out for TMP, so it was never measured quantitatively in any analysis.

4 EXPERIMENTAL PROCEDURES

4.1 GENERAL OVERVIEW

The objective of the TTPE program was to demonstrate that the analytical methods described here are capable of quantitatively measuring GB and GD in specified test matrices at the concentration levels identified in Table 3. To accomplish this objective, the test matrices were spiked with predetermined concentrations of agent and extracted with solvent. The agent present in each extract was then preconcentrated onto a sorbent cartridge, thermally desorbed and trapped onto a focusing tube, thermally desorbed onto a gas-chromatographic column, and detected by using a flame photometric detector. The experimental procedures involved in each of these steps are described in this section.

4.2 CHEMICAL AGENTS

Isopropyl methylphosphonofluoridate (GB) and pinacolyl methylphosphonofluoridate (GD) were received as Chemical Agent Standard Analytical Reference Material (CASARM) from the Army's Chemical Transfer Facility at Edgewood, Maryland, and stored within the Chemical Repository at DPG. Surety-certified personnel from DPG prepared both the agent calibration standard solutions and the matrix spiking solutions. The resulting "dilute"-level stock solutions of each agent were prepared from CASARM stocks of neat agent according to DPG standard operating procedure (SOP) No. MT-C-7-13 and were then transferred to the analytical laboratory for storage and use.

4.3 PREPARATION OF STANDARD SOLUTIONS

4.3.1 Calibration and Check Standards

A stock standard calibration solution was prepared for each agent in chloroform. The concentration of this stock solution was verified by independent analysis prior to use. The concentration value resulting from this analysis was required to be within 15% of the target solution concentration, or a new stock solution was prepared. A set of working calibration standard solutions was prepared in chloroform from this dilute stock solution.

The set of working calibration standards for each agent consisted of eight solutions. The concentrations of these solutions were determined so that 10- μ L spikes onto sorbent cartridges according to the procedures discussed below would deliver the desired amounts of agent to the GD or GB analytical system. The desired amounts corresponded to the following specified

fractions of the TRL: 0×TRL, 0.375×TRL, 0.5×TRL, 1×TRL, 2×TRL, 5×TRL, 10×TRL, and 12.5×TRL. By selecting the proper spiking concentrations, solvent extraction volume, and volume of extract delivered to the sorbent cartridge, the same set of calibration standard solutions could be used to calibrate the GC/FPD system regardless of the test matrix. Calibration was carried out by spiking 10-μL aliquots of the working calibration standards onto sorbent cartridges and analyzing these cartridges as samples. Table 7 lists the concentrations of the working calibration standard solutions for GB and GD.

In addition to the set of working calibration standards, a check standard was prepared at a concentration corresponding to 8×TRL. The check standard was prepared from a different lot of neat CASARM by surety-certified personnel other than those who prepared the stock calibration solutions. The check standard was run each time the calibration standards were run and during each method certification run for a given agent. The check standard was run by spiking 10 μL of check standard solution onto a sorbent cartridge and analyzing the tube in the usual way. The concentrations of the check standard solutions used for GB and GD are also given in Table 7.

4.3.2 Matrix Spiking Solutions

Table 8 lists the concentrations of the agents in the working spiking solutions. The same spiking solutions were used for all matrices. The solvent used for the spiking solutions was 2-propanol instead of chloroform in order to minimize evaporative changes in standard concentrations and for its compatibility with the aqueous HTH decon fluid used as one of the test matrices. Samples of each matrix were spiked at the indicated multiples of their respective TRLs by applying 10-μL aliquots of the spiking solutions to 10-g samples of the metal and soil matrices and to 20-mL samples of the decon fluids.

TABLE 7 Calibration and Check Standard Solutions for GB and GD^a

Standard Type	Concentration Level (Multiple of TRL)	GB Standard Concentration (ng/μL)	GD Standard Concentration (ng/μL)
Calibration	0×TRL	0.0	0.0
Calibration	0.375×TRL	0.18	0.045
Calibration	0.5×TRL	0.24	0.060
Calibration	1.0×TRL	0.48	0.12
Calibration	2.0×TRL	0.96	0.24
Calibration	5.0×TRL	2.4	0.60
Calibration	10.0×TRL	4.8	1.2
Calibration	12.5×TRL	6.0	1.5
Check	8.0×TRL	3.84	0.96

^a Note: 10 μL of each solution were used per spike.

Concentrations of spiking solutions for the decon solution matrix were calculated so that a 10- μ L aliquot of the spiking solution applied to a 20-mL aliquot of the HTH decon fluid matrix will achieve half the indicated fractions of the TRL. The factor of one-half accounts for the 1:2 dilution that occurs in reducing the hypochlorite ion and making necessary adjustments in pH starting from the original decon solution samples, as discussed in Section 4.4.1.

As discussed in Section 3.2.4, in addition to being spiked with agent, each sample was also independently spiked with a solution of TMP in isopropanol. The TMP solution concentration was chosen so that the number of moles of TMP spiked into each sample was equal to the number of moles of agent spiked into a sample at the 1 $\times$ TRL level. The TMP spiking level was the same for all samples spiked with a given agent and was used to qualitatively evaluate the efficiency of the solvent extraction step.

4.4 TEST MATRICES

4.4.1 5% HTH Decon Fluids

Preparation of NC and SC Decon Fluids

HTH decon fluid was prepared by adding water to commercially available solid HTH, which is approximately 65% calcium hypochlorite and 35% inert ingredients. One-liter batches of aqueous 5% HTH solution were prepared as needed. A 5% HTH solution contains approximately 3.25% calcium hypochlorite.

TABLE 8 Matrix Spiking Solutions for GB and GD^a

Concentration Level (Multiple of TRL)	GB Spiking Solution Concentration (ng/ μ L)	GD Spiking Solution Concentration (ng/ μ L)
0 $\times$ TRL	0.0	0.0
0.5 $\times$ TRL	12.0	3.0
1.0 $\times$ TRL	24.0	6.0
2.0 $\times$ TRL	48.0	12.0
5.0 $\times$ TRL	120.0	30.0
10.0 $\times$ TRL	240.0	60.0

^a Note: 10 μ L of each solution were used per spike.

To make the 5% HTH solution less hostile to the chemical agents, the hypochlorite present in the solution was reduced by the addition of ascorbic acid, which reacts with hypochlorite to form chloride ion, water, and other products. To determine the amount of ascorbic acid required to neutralize a given amount of a 5% HTH solution, a 50-mL aliquot of HTH solution was titrated potentiometrically with one-molar ascorbic acid. From the result of this titration, the total amount of ascorbic acid required to reduce the desired amount of 5% HTH solution was computed and increased by 10% to ensure complete neutralization of the hypochlorite ion. Following the reduction step, 2N NaOH was used to adjust the pH of the reduced decon solution to 5 ± 0.1 in order to minimize hydrolysis of the GB or GD. The quantity of caustic required to adjust the 50-mL aliquot of reduced HTH solution to the proper pH was also recorded.

Following the reduction and pH adjustment of the 50-mL aliquot, a fresh 500-mL sample of the same 5% HTH solution was placed in a 1-L flask, and the calculated quantities of ascorbic acid and 2N NaOH were added for this volume of sample. This solution was then diluted to volume (1 L) with a saturated aqueous solution of NaCl. The use of a NaCl solution serves to reduce the solubility of the GB or GD in the aqueous matrix and hence facilitate its extraction. If necessary to achieve homogeneity, the solution was allowed to settle overnight and then the supernatant was decanted, leaving approximately 5% of the total volume behind as undissolved material. The resulting solution is referred to as the non-contaminated-filtered, reduced, pH-adjusted, and diluted (NC-FRPD) solution. The pH of the solution was checked each time a sample was removed for performing a set of spiking experiments and adjusted to the proper pH (5 ± 0.1), if necessary.

Batches of SC decon fluid were prepared as needed by DPG surety-certified personnel. To prepare SC decon fluids, 12.5 g GB was added to 1,300 mL of 5% HTH solution, and 12.5 g GD was added to 1,100 mL of 5% HTH solution. The agent was destroyed by reaction with the hypochlorite, generating several possible degradation products. One agent was used in each batch. As a result of this procedure, any SC decon fluid used was known to have been exposed to the agent of interest prior to the analytical method demonstration experiments.

Once prepared, this material was subjected to the same hypochlorite ion reduction, pH-adjustment, and dilution procedures described above for the NC decon fluid except that the sediment was allowed to remain in the decon fluid. The resulting solutions, one using GB and one using GD, are referred to as the suspect contaminated-filtered, reduced, pH-adjusted, and diluted (SC-FRPD) solutions.

Procedures for Spiking and Extracting NC or SC Decon Fluid

A 20-mL aliquot of FRPD solution adjusted to the appropriate pH was placed in a 40-mL conical glass centrifuge tube. Using a 25- μ L syringe, a 10- μ L aliquot of surrogate spiking solution, followed by a 10- μ L aliquot of the desired agent spiking solution, was dispensed into the FRPD solution. The centrifuge tube containing the spiked solution was immediately capped

and the solution was swirled for 2–5 s. The extraction was then performed immediately by adding 2 mL of chloroform, capping the centrifuge tube, vortexing (i.e., mixing thoroughly using an electrically powered “touch mixer”) for 30 s, and then placing it in the centrifuge at 2,000 rotations per minute (rpm) for 2 min to facilitate separating the chloroform extraction layer at the bottom of the centrifuge tube. A portion of the chloroform layer was carefully transferred into a prelabeled 2-mL sample vial using a Pasteur pipette. A 40- μ L aliquot of the chloroform extract was spiked onto the sorbent cartridge through a glass fuzz tube as described in Section 4.5, and the sorbent cartridge was analyzed as described in Section 4.6.

4.4.2 Metal Substrate

Preparation

The circular metal disks used for this study were prepared by the machine shop at DPG from 20-gauge, type 316 stainless steel, with a diameter of 0.25 in. and an average weight of 0.180 g each. The disks had nonporous, nonpermeable surfaces free of paint, oil, and other surface contamination. The disks were washed by soaking overnight in a mild detergent solution. After initial washing, the disks were rinsed five times with deionized water, washed three times with methanol, then washed three times with hexane. The disks were then air-dried on aluminum foil for 2 h to remove hexane and then oven-dried at 110°C. Once they had been cleaned, the metal disks were only handled by using tweezers. In all instances, the metal disks were used once and then discarded.

Procedures for Spiking and Extracting Metal Disks

Approximately 10 g of metal disks were placed in a 15-mL glass centrifuge tube. Analysts used 25- μ L syringes to spike the surface of the disks with a 10- μ L aliquot of the surrogate spiking solution followed by a 10- μ L aliquot of the agent spiking solution. Immediately after the metal disks were spiked, exactly 7.5 mL of chloroform was pipetted into the centrifuge tube, the tube was capped, and the solution was mixed by vortexing for 30 s. A portion of the extract was then transferred using a Pasteur pipette into a properly labeled GC vial. A 150- μ L aliquot of this chloroform extract was spiked onto a sorbent cartridge, as described in Section 4.5, and the cartridge was analyzed as described in Section 4.6.

4.4.3 Soil

Selection and Preparation of NC and SC Soil Samples

DPG personnel prepared a composite of representative DPG soils (DPG standard soil) with no known history of exposure to chemical agents; this sample was used as the NC soil matrix. The SC soil matrix was collected by DPG personnel from a SWMU area at DPG in which HD contamination is suspected. Both soil matrices were used “as is,” without drying. The soil was screened to remove large pebbles, stones, and debris, and the larger soil clumps were reduced in size by crushing.

Procedure for Spiking and Extracting Soil Samples

Approximately 10 g of soil (weight recorded to the nearest 0.1 g) was weighed into a 40-mL centrifuge tube. Analysts used 25- μ L syringes to spike the soil with a 10- μ L aliquot of the surrogate spiking solution and a 10- μ L aliquot of the desired agent spiking solution. After spiking, exactly 7.5 mL of chloroform was pipetted into the centrifuge tube, the tube was capped, and the solution was mixed by vortexing for 30 s. The centrifuge tube was then rotated at 2,000 rpm for 2 min to separate the chloroform extract layer. A portion of the extract was then transferred using a Pasteur pipette into a prelabeled GC vial. A 150- μ L aliquot of the extract was spiked onto a sorbent cartridge as described in Section 4.5, and the tube was analyzed as described in Section 4.6.

4.5 PROCEDURE FOR SPIKING SORBENT CARTRIDGES

The Perkin Elmer ATD 400 thermal desorption unit used in this work uses sorbent cartridges measuring 0.25 in. outside diameter by 3.5 in. long. The cartridges are obtained from a commercial supplier and packed with 60–80-mesh Tenax TA resin, which resides between the head of the cartridge containing a section of glass frit and the back of the cartridge containing silanized glass wool and a metal spring clip to keep the sorbent resin in the cartridge.

The apparatus used to preconcentrate solutions for GB and GD analysis is illustrated schematically in Figure 4. Figure 4(a) shows the “fuzz” tube, Figure 4(b) the sorbent cartridge, and Figure 4(c) the assembled apparatus. The sorbent cartridge resin (Tenax TA[®] resin), on which the agents are preconcentrated, is in the downstream tube (the sorbent cartridge), while the upstream tube, called the “fuzz” tube, contains silanized glass wool. The purpose of the fuzz tube was to provide a large surface area from which the solvent and analyte could volatilize to ensure that only vapor enters the sorbent cartridge. Entry of liquid solvent into the sorbent cartridge, which is designed primarily for gas-phase sample collection, may result in physical damage to

FIGURE 4 Sorbent Cartridge Spiking Apparatus

the resin or in solvent washing of the analyte from the sorbent resin. Each fuzz tube was used once and the glass wool removed and discarded. The tube was then cleaned with chloroform, acetone, and hexane, oven dried, re-silanized if necessary, and then repacked with silanized glass wool.

The fuzz tube and sorbent cartridge were connected by using a standard Teflon® compression fitting. After assembly, room air was passed through the system at a rate of up to 500 mL/min by applying vacuum to the downstream end. An aliquot of the agent-containing chloroform extract was injected onto the silanized glass wool in the fuzz tube. The amount of extract used depended on the matrix and varied from 40 µL to 150 µL. The air stream evaporated the solvent and the agent from the surface of the glass wool and transferred the vapors to the cartridge, where the agent was retained on the resin. The solvent was not retained. After the extract was transferred to the fuzz tube, an additional 40 µL of chloroform was injected onto the fuzz tube in order to wash the tube and further promote the transfer of any residual agent to the cartridge. After allowing sufficient time for transfer of the agent onto the cartridge (approximately 5 min), the analyst stopped the air stream, capped both ends of the cartridge, and retained the cartridge for analysis. When sorbent cartridge preconcentration was completed, GB and GD were quantitatively transferred from the solvent extract injected into the fuzz tube to the resin in the sorbent cartridge.

4.6 SORBENT CARTRIDGE ANALYSIS BY THERMAL DESORPTION AND GAS CHROMATOGRAPHY

4.6.1 Thermal Desorption

Each GC was equipped with a Perkin Elmer Model ATD-400 Automated Thermal Desorption (ATD) unit. The ATD unit is an automated sample introduction system that contains a carousel allowing the sequential analysis of up to 50 sorbent cartridges. The unit automatically checks sorbent cartridges for leaks and thermally desorbs, traps, and transfers the analyte to the GC column in a preprogrammed sequence. The procedure used for preconcentrating sample extracts, calibration solutions, or surrogate spiking solutions using sorbent cartridges was described in Section 4.5.

Analysis of the cartridges on which an extract or an agent solution standard had been preconcentrated began by thermally desorbing the agent from the cartridge onto the GC column using the ATD unit. The sorbent cartridge caps were first removed automatically from the ends of the tube. A stream of helium was then passed through the tube and heat was applied, causing the agent to be desorbed from the resin into the helium stream. The helium carried the agent to a much smaller sorbent cartridge, also packed with Tenax TA® and known as the focusing tube, which was cooled to 0°C using Peltier cooling. Cooling the focusing tube ensured that the analyte was completely trapped. Once the agent had been transferred onto the focusing tube, the valving was changed so that the helium stream passing through the focusing tube was directed to

the GC column. The focusing cartridge was flash-heated, and the agent desorbed in a compact pulse that was delivered through a heated transfer line to the GC column. All the steps described for this procedure were performed automatically by preselecting the desired conditions and entering the appropriate parameter values associated with each step into the Perkin-Elmer TURBOCHROM® software.

In this study, the GC column was threaded through the ATD-GC transfer line and connected directly to the switching valve within the ATD instead of using a standard uncoated capillary transfer line between the ATD and the GC. The temperature of the transfer line was maintained at the initial temperature for the GC column temperature program (60°C).

Table 9 provides a typical sample sequence for an instrument calibration or method precertification run. A typical sample sequence for a method certification run is provided in Table 10.

By eliminating the solvent in the sorbent cartridge preconcentration process, volumes of solvent extract or standard solutions much greater than the typical 1–2- μ L direct liquid injection volume can be analyzed. This process results in a considerable increase in method sensitivity over the more common direct liquid injection technique. An additional advantage of the sorbent cartridge preconcentration and thermal desorption methodology is that many of the polar and nonvolatile substances that may be present in the solvent extracts are not readily volatilized from either the fuzz tube or the sorbent cartridge. So the method appreciably reduces chromatographic interferences that may arise from decomposition of these substances in the injection port or on the column when conventional liquid injection methods are used. However, as with any preconcentration technique, volatile interfering components in the sample extract may also be preconcentrated and may compromise the analyte detection or separation.

TABLE 9 Typical Calibration Run Sample Sequence

Tube No.	Calibration Standard	Description
1	0 $\times$ TRL	Sorbent cartridge spiked with 10 μ L chloroform
2	0.375 $\times$ TRL	Sorbent cartridge spiked at designated agent level
3	0.5 $\times$ TRL	Sorbent cartridge spiked at designated agent level
4	1.0 $\times$ TRL	Sorbent cartridge spiked at designated agent level
5	2.0 $\times$ TRL	Sorbent cartridge spiked at designated agent level
6	5.0 $\times$ TRL	Sorbent cartridge spiked at designated agent level
7	10.0 $\times$ TRL	Sorbent cartridge spiked at designated agent level
8	12.5 $\times$ TRL	Sorbent cartridge spiked at designated agent level
9	12.5 $\times$ TRL	Sorbent cartridge spiked at designated agent level
10	0 $\times$ TRL	Sorbent cartridge spiked with 10 μ L chloroform
11	8 $\times$ TRL	Sorbent cartridge spiked with check standard, independently prepared
12	Blank	Sorbent cartridge analyzed unspiked

TABLE 10 Typical Method Certification Run Sample Sequence

Tube No.	Spiking Solution	Description
1	Solvent control	2 mL chloroform spiked with 1×TRL of agent and 1×TRL of surrogate (TMP)
2	0×TRL	Matrix spiked with 10 µL of 2-propanol
3	0×TRL	Matrix spiked with 10 µL of 2-propanol
4	0.5×TRL	Matrix spiked at designated agent level
5	1.0×TRL	Matrix spiked at designated agent level
6	2.0×TRL	Matrix spiked at designated agent level
7	5.0×TRL	Matrix spiked at designated agent level
8	10.0×TRL	Matrix spiked at designated agent level
9	10.0×TRL	Matrix spiked at designated agent level
10	0.375×TRL	Sorbent cartridge spiked at designated agent level (low calibration standard)
11	12.5×TRL	Sorbent cartridge spiked at designated agent level (high calibration standard)
12	8.0×TRL	Sorbent cartridge spiked with check standard, independently prepared
13	Solvent control	2 mL chloroform spiked with 1×TRL of agent and 1×TRL of surrogate (TMP)

4.6.2 Gas Chromatography

Gas chromatography was used for analysis of sample extracts. Analyses were performed using two Perkin Elmer Model 9000 Autosystem GCs equipped with ATD-400 thermal desorption units as described above. One GC system was dedicated to GB analyses and the second system to GD analyses. Both GCs were equipped with FPDs, which were operated in the phosphorus mode. A third ATD-400 unit was employed for reconditioning used sorbent cartridges. The three GC systems were networked to two computerized data acquisition stations loaded with Perkin-Elmer's TURBOCHROM® software system, which was used both to control the GC systems and to process the chromatographic data.

Table 11 gives the GC operating parameters for both GB and GD analyses conducted as described above.

4.7 CONFIRMATION METHODS

GC/MS confirmation methods were developed for GB and GD. The purpose of confirmation is to avoid false positive identifications and to provide additional data to confirm positive hits. For both agents, the sample extract is introduced into the GC/MS by using the Dynatherm ACEM 900 thermal desorption system configured to a Hewlett-Packard GC equipped with a mass-selective detector (MSD). Confirmation is accomplished by comparing the sample

mass spectrum obtained by using the MSD in the scan mode with entries in a mass spectral library of standard compounds. The Hewlett-Packard GC/MS Chem Station software can automatically make this comparison and assign a "quality of hit" value. Ion ratios can also be examined manually. The analyst must ultimately make the final decision regarding the identification of the substance in question on the basis of all the available data.

TABLE 11 Gas Chromatographic Operating Parameters^a

GC Operating Parameter	Description
Column	30-m × 0.8 mm outside-diameter (OD), DB-5 bonded phase, 0.53 mm inside-diameter (ID), 1.5 µm film thickness
Column temperature program	Ramp from 60°C to 90°C at 10°C/min; hold at 90°C for 1 min; ramp from 90°C to 220°C at 45°C/min; hold at 220°C for 1.1 min
Detector temperature (FPD/phosphorus mode)	250°C
Sorbent cartridge desorption temperature	250°C
Sorbent cartridge desorption time	5 min
Focusing trap temperature (low)	0°C
Focusing trap temperature (high)	275°C
Focusing trap purge time	4 min
Transfer line temperature	60°C ^b
Carrier gas	Ultrapure helium
Carrier gas flow rate	20 mL/min ^c
ATD valve temperature	200°C
Desorption flow rate	80–100 mL/min

^a The description is common to both GB and GD analyses.

^b Same as the initial column temperature.

^c Not less than 10 mL/min.

5 RESULTS AND DISCUSSION

The results of the method validation work are presented and discussed in this section. Section 5.1 discusses the GB and GD results in relation to the DQOs described in Section 2.4. Section 5.2 contains discussions of several additional issues, including the treatment of outliers, the dependence of the method standard deviation on spiking level, and others. Section 5.3 summarizes the QC sample results.

5.1 GB AND GD METHOD VALIDATION RESULTS

5.1.1 Attainment of Data Quality Objectives

Prior to the method certification experiments, the calibration model for each agent was successfully precertified by following the procedure described in Section 3.1.1. Following precertification, a minimum of four runs were carried out for each agent/matrix combination by using matrix samples spiked at the levels specified by the USATHAMA QAPP (as described in Section 3.1). Some runs were rejected because of QC problems, instrument problems, or the presence of outliers. These runs were repeated. For each agent/matrix combination, the statistical analyses required for method certification were based on four sets of certification data.

On several occasions, one or more outliers were found; depending on the circumstances, the certification run or the runs involved were repeated or the outlier was replaced with a value computed as described in Section 5.2.1. The substitution of an outlier with a computed replacement value was required so that an equal number of "measured" concentrations were available for each target spiking level. When no outliers were identified in a dataset, the linearity and zero-intercept tests, as well as the computation of the CRL, were completed by using both the IRPQAP software and a spreadsheet designed to perform the necessary calculations. When an outlier had been replaced, only the spreadsheet was used. The final certification data for each agent/matrix combination contained no more than one outlier replacement value and, as shown below, satisfied the QAPP requirements for linearity and zero intercept by wide margins, meeting the associated DQOs.

The linearity and zero-intercept checks were conducted on the basis of a linear regression of found agent concentrations against target spiked values. The quality of the linear regressions of found against true agent concentrations is shown by the values listed in Table 12, which summarizes the results of the regression analyses. The regression analyses were carried out by using the found and target amounts of agent used (in nanograms) rather than the actual agent concentrations in the matrix. The concentrations are directly proportional to the agent amounts; the results of the statistical tests are not altered. The only parameter value affected by this choice is the intercept. For each agent/matrix combination, the slope and intercept of the regression line

TABLE 12 Summary of Linear Regression Results for GB and GD Found versus True Amounts

Agent	Matrix ^a	Slope ^{b,c}	Intercept ^{c,d}	(R ²) ^b	Degrees of Freedom
GB	NCM	1.156 ± 0.025	0.131 ± 0.561	0.9976	22
	NCDF	1.087 ± 0.013	0.075 ± 0.286	0.9993	21
	SCDF	1.031 ± 0.026	-0.034 ± 0.579	0.9968	22
	NCS	1.053 ± 0.022	-0.065 ± 0.499	0.9977	21
	SCS	1.120 ± 0.030	0.265 ± 0.677	0.9963	21
GD	NCM	1.139 ± 0.011	-0.012 ± 0.063	0.9995	21
	NCDF	1.000 ± 0.030	0.052 ± 0.169	0.9954	22
	SCDF	1.055 ± 0.027	0.045 ± 0.152	0.9966	21
	NCS	1.174 ± 0.030	-0.003 ± 0.170	0.9966	22
	SCS	1.078 ± 0.023	0.053 ± 0.129	0.9977	21

^a NCDF = noncontaminated decon fluid, SCDF = suspect-contaminated decon fluid, NCS = noncontaminated soil, SCS = suspect-contaminated soil, NCM = noncontaminated metal.

^b Dimensionless.

^c The indicated uncertainties represent 95% confidence limits.

^d Units = nanograms.

and the associated 95% confidence limits are shown along with the value of R² for the regression and the associated number of degrees of freedom. R² is a measure of the extent to which a linear relationship explains the experimental data; R² equal to one indicates perfect agreement. The number of degrees of freedom represents the number of measured values (24; i.e., six target values with four replicates each) minus the number of parameters that were estimated in the regression analysis (2) minus the number of outliers replaced. Thus, the number of degrees of freedom is 22 if no outliers were found and 21 if one outlier was replaced. The regression results were obtained using data that included outlier replacement values, but no significant differences would be seen in the results if the outliers had been simply deleted from each dataset.

All but one of the slopes listed in Table 12 are significantly higher than the ideal value of unity at the 95% confidence level. These results indicate that the analyses produced concentration values that were systematically higher than the true values. The reason for this bias is not known. Additional related results are presented and discussed in Section 5.2.3 and Appendix C. The 95% confidence limits for all intercept values include the value zero by wide margins, and it is clear that no systematic bias or background exists at zero agent concentration. Finally, all R² values exceed 0.995, and two exceed 0.999. Thus, the linearity of the found versus true concentration

curve is excellent for both GB and GD and for all matrices considered. Summaries of the certification data for GB and GD are given in Appendix B.

Table 13 presents a summary of the numerical measures of method validity associated with the data quality objectives discussed in Section 2.4. Also shown is the TRL for each agent/matrix combination.

5.1.2 Certified Reporting Limit

As shown in Table 13 (columns 3 and 4), all CRL values are significantly lower than the corresponding TRLs. Based on Section 14 of the USATHAMA QAPP, the CRL is that true concentration at which there is less than a 5% chance of obtaining a found concentration below that value which can just be distinguished from zero at the 95% confidence level. The CRL represents one possible definition of a quantitation limit, the true concentration at or above which the method can be expected to provide quantitative results. Based on the results listed in Table 13, the analytical methods described here are clearly capable of producing quantitative

TABLE 13 Summary of Method Development and Validation Results

Agent	Matrix ^a	TRL ^b	CRL ^b	MDL ^b	S/TRL ^c	Total Error (1) ^d	Total Error (2) ^d
GB	NCM	24	15	8.2	0.190	49.3	71.8
	NCDF	24	8.3	2.3	0.023	11.3	14.1
	SCDF	24	18	3.7	0.035	7.6	11.8
	NCS	24	15	4.4	0.025	14.6	17.5
	SCS	24	19	1.7	0.078	34.1	43.3
GD	NCM	6	1.8	0.39	0.052	19.6	25.8
	NCDF	6	5.3	0.96	0.073	24.0	32.6
	SCDF	6	4.5	0.49	0.046	16.6	22.0
	NCS	6	4.6	1.2	0.079	34.1	43.4
	SCS	6	3.8	0.57	0.031	16.6	20.2

^a NCDF = noncontaminated decon fluid, SCDF = suspect-contaminated decon fluid, NCS = noncontaminated soil, SCS = suspect-contaminated soil, NCM = noncontaminated metal.

^b Units are ng/mL (ppb) for decon fluid, ng/g (ppb) for soil and unpainted metal.

^c Method standard deviation (S at true concentration = TRL) divided by TRL; dimensionless.

^d Total error (1) is based on Equation (2) with factor = 2.0; total error (2) uses Equation (2) with factor = 3.182; units are percentage (parts per hundred); total error values are evaluated at the 1×TRL spiking level.

results for GB and GD in the target concentration ranges for unpainted metal, soil, and spent HTH decon fluid matrices. The related DQO, that the CRL should be less than the TRL, is met for all agent/matrix combinations. The differences in and variability of the CRL values presented in Table 13 are discussed in Section 5.2.3.

The USATHAMA QAPP describes the CRL but does not give a formula for its computation. As part of the data analysis in this study, the following equation was derived:

$$CRL = t_{\alpha} \left(\frac{s_{y,x}}{b} \right) [f(0) + f(CRL)] \quad (4)$$

This equation is based on a standard linear regression of found versus target concentrations. In Equation (4), $s_{y,x}$ denotes the mean-square deviation of found concentrations about the regression line; b denotes the slope of the regression line; and t_{α} denotes the two-tailed, 90% confidence level critical value of the Student's-t distribution with $n-2$ degrees of freedom (n is the total number of experimentally measured values used in the method certification process, including blanks). The function f appearing in Equation (4) is given by the following:

$$f(X) = \left[1 + \frac{1}{n} + \frac{(X - \bar{X})^2}{\sum (X_k - \bar{X})^2} \right]^{1/2} \quad (5)$$

where the summation is over all target concentrations used, and $\bar{X}$ denotes the overall mean of these target values. Equation (4) for the CRL incorporates the assumption that the standard deviation of measured values is independent of the spiking level.

Equation (4) was used to compute the CRL in the spreadsheet designed to duplicate and check the IRPQAP computations. Because the equation is an implicit rather than an explicit expression for the CRL, an iterative method was used to compute the CRL value. The function f is slowly varying; an iterative method starting with an initial estimate in which CRL is set to zero on the right side gives CRL values accurate to six significant figures in just 2-3 iterations. The values obtained using this procedure differed by no more than 0.5% from those computed by the IRPQAP software.

The CRLs listed in Table 13 clearly show a difference between GB and GD. This difference can be attributed to the use of a different set of spiking levels for the two agents associated with the different TRLs used, combined with systematically lower standard deviations at lower spiking levels for both agents. This dependence of standard deviation on spiking level is a common observation in trace chemical analysis in general. As explained in Section 3.1.2, the spiking levels are defined by the QAPP in terms of the TRL value, and because the TRL for GD is one fourth the TRL for GB, the corresponding spiking levels were also lower by the same factor. As a result, the standard deviation about regression $s_{y,x}$ appearing in Equation (4) is lower for GD than that for GB, and the computed CRL is also reduced by approximately the same factor.

However, outside of these obvious and explainable differences, no clear evidence of any significant difference in method performance between matrices or between agents GB and GD is apparent in the results. There are at least two reasons for this finding. One is that, after the sample extraction step, the analytical methods are virtually identical; they involve the same preconcentration, thermal desorption, and GC stages using nearly identical equipment and operating conditions. Also, the extractions were performed as soon as possible after spiking the samples to ensure that the amount of agent present was known with the greatest degree of accuracy. The second reason is that GB and GD are very similar chemically and are expected to behave in similar ways. Thus, the CRL values show no clear dependence on agent (between GB and GD only) or matrix type beyond that caused by the variation of method standard deviation with spiking level.

The CRL is computed from the set of experimental results obtained from samples spiked at specified levels. The experimental results are subject to random statistical fluctuation, and so therefore is the CRL value. If the methods for a given agent are all essentially equivalent, the same CRL should be obtained for each matrix within statistical fluctuation, and the variability in the values listed in Table 13 should indicate the level of accuracy with which the CRL can be found. The standard deviation of the five CRL values divided by the mean CRL value equals 0.28 for GB and 0.33 for GD.

Two other indicators are available to judge the accuracy of the CRL. Using the Jackknife technique (Mosteller and Tukey 1977), the relative standard deviation (RSD) of the CRL was evaluated for the GB-SCDF dataset to be 0.25. Also, a Monte Carlo calculation was carried out based on a model system in which no systematic error is present at any spiking level but in which the standard deviation of measurements increases linearly with spiking level at a rate typical of what was observed in this study. The RSD for the model CRL computation was found to be 0.27 in this calculation. The general agreement between the RSD of measured GB and GD CRL values and the other two estimates of the CRL RSD supports the conclusion that the CRL values are essentially independent of the agent and matrix type. On the basis of these three independent pieces of evidence, it appears that the CRL for these analytical methods can be found to within about 30% with a single determination using four certification runs per dataset.

Table 13 also provides the MDLs (column 5), defined by EPA as "the minimum concentration of a substance that can be measured and reported with 99% confidence that the analyte concentration is greater than zero and is determined from analysis of a sample in a given matrix type containing the analyte" (EPA 1986). All MDL values are substantially less than the associated CRLs. The MDL was computed as the one-tailed Student's t-statistic for three degrees of freedom and 99% confidence (value = 4.541) times the standard deviation of the four measured concentration values at the 0.5×TRL spiking level (EPA 1986). If an outlier had been replaced in the set of values for this spiking level, the appropriate t-statistic for two degrees of freedom (value = 6.965) was used instead and only the three actual measured values were used in computing the standard deviation. The MDL for GB-NCM is unusually large compared with the other GB values because of an unusually large method standard deviation.

The MDLs for GB and GD also exhibit a significant systematic difference, again caused by the combination of the lower spiking levels used for GD and the systematic increase of method standard deviation with spiking level. Because the lowest spiking level for GD was one fourth of that for GB, the standard deviation of measurements at the lowest spiking level for GD, the basis for computing the MDL, was lower than that for GB.

5.1.3 Method Standard Deviation and Total Error

The sixth column of Table 13 lists the values of the method standard deviation (S) divided by the TRL; the values in this table are ratios evaluated at the $1 \times \text{TRL}$ spiking level. As described above, the standard deviation (S) was computed as the standard deviation of the measured values only; no outlier replacement values were used. The associated DQO (described in Section 2.4) was that the ratio of the method standard deviation to the true concentration should be less than 0.1 for all agent/matrix combinations and for spiking levels near the TRL. The significance of the ratio S (at the $1 \times \text{TRL}$ level)/TRL is that the probability of false positives, given the use of the Tolerance Level as discussed in Section 2.3, is expected to be acceptably low for S/TRL values below approximately 0.1. As shown in Table 13, the value of S/TRL at the TRL is substantially less than 0.1 for all agent/matrix combinations but one. Again, the GB-NCM value (0.190) is anomalous. The value 0.190 is the highest value of S/(true concentration) for any agent/matrix combination and any spiking level in the combined GB and GD dataset. GD and GB have similar chemical properties, and if the NCM matrix presented some special problem leading to high method standard deviations, the problem should also appear in the remainder of the GB-NCM data as well as in the GD-NCM data. However, three of the five GB values and the GD-NCM values are entirely acceptable. In view of these observations, the high values of the MDL and of S/TRL for GB-NCM are believed to be caused by some unidentified problem in spiking that was not repeated in the other datasets.

The Total Error results (columns 7 and 8 in Table 13) are based on Equation (2) in Section 2.4, and are given for both values of the factor t discussed there. In Table 13, Total Error values are evaluated only for the $1 \times \text{TRL}$ spiking level. The t -value 2.0 was used in the original work (McFarren, et al. 1970) and has gained broad acceptance when the number of measurements used is reasonably large; however, the value 3.182 is more appropriate here because the method standard deviation and bias at any spiking level are based on only four datapoints. McFarren, et al. characterized analytical methods as "excellent" if the Total Error is less than 25%, "acceptable" if the Total Error lies between 25% and 50%, and "unacceptable" if the Total Error exceeds 50%. Using the original definition, seven of the ten agent/matrix combinations have Total Errors in the excellent range, three in the acceptable range, and none are unacceptable. Using the definition of Total Error incorporating the factor 3.182, five of the ten values are still in the excellent range; four are acceptable; and one, the GB-NCM value, is unacceptable.

In order to put the method standard deviation and Total Error results into perspective, it is useful to present those data for all spiking levels, not just for the 1×TRL level. Tables 14 and 15 show the ratio of the method standard deviation to the true concentration (the RSD) for the GB and GD data, respectively, at all spiking levels. The GB values are more or less randomly distributed about a mean of 0.050 with a standard deviation of 0.023; the first two GB-NCM datapoints (0.190 and 0.151) were omitted from the analysis because they are clearly anomalous. The GB-NCDF data are systematically lower than this average, otherwise no clearly significant variations are seen in these results. For GD, the situation is reversed; the GD-NCDF data are somewhat high and the GD-NCM data relatively low. The GD values are distributed about a mean of 0.052 with a standard deviation of 0.024, almost identical to the corresponding GB numbers. The methods used to analyze GB and GD in different matrices were all very similar, and the general similarity of the relative method standard deviation results is not unexpected. If these data are considered as a whole, it appears that the expected RSD for the G-agent methods is close to 5%, independent of the matrix. As mentioned above, the high values for the first two GB-NCM spiking levels are believed to indicate some temporary difficulty in the laboratory operations rather than a basic or significant difference of the GB-NCM results from all the other data. The associated DQO is considered to be achieved to an acceptable degree, on the basis of these results.

TABLE 14 Ratio of Method Standard Deviation to True Concentration: GB Data^a

Spiking Level	NCM	NCDF	SCDF	NCS	SCS
0.5×TRL	0.151	0.041	0.068	0.089	0.039
1.0×TRL	0.190	0.023	0.035	0.025	0.078
2.0×TRL	0.084	0.021	0.026	0.061	0.086
5.0×TRL	0.054	0.047	0.086	0.052	0.055
10.0×TRL	0.038	0.012	0.037	0.032	0.055

^a Values are dimensionless.

TABLE 15 Ratio of Method Standard Deviation to True Concentration: GD Data^a

Spiking Level	NCM	NCDF	SCDF	NCS	SCS
0.5×TRL	0.029	0.071	0.036	0.092	0.042
1.0×TRL	0.052	0.073	0.046	0.079	0.031
2.0×TRL	0.009	0.103	0.090	0.059	0.042
5.0×TRL	0.021	0.041	0.050	0.063	0.031
10.0×TRL	0.021	0.059	0.050	0.057	0.044

^a Values are dimensionless.

Tables 16 and 17 show the Total Error values obtained for each agent/matrix combination and each spiking level. The tables present Total Error values for each of the two values considered for the parameter t (2.0 and 3.182). Using a value of 2.0, we found 19 Total Error values for GB in the excellent range and 6 in the acceptable range, with no unacceptable values. Of the six values in the acceptable range, from 25% to 50%, four were found in the NCM data and two in the SCS data. If the value 3.182 is used, 12 Total Error values for GB lie in the excellent range, 11 in the acceptable range, and 2 in the unacceptable range. The two unacceptable values correspond to the first two anomalous GB-NCM datapoints, which have already been discussed. The associated Total Error DQO is considered to have been achieved to a satisfactory degree.

TABLE 16 Total Error Values: GB Data^a

Spiking Level	NCM		NCDF		SCDF		NCS		SCS	
	(2.0)	(3.182)	(2.0)	(3.182)	(2.0)	(3.182)	(2.0)	(3.182)	(2.0)	(3.182)
0.5×TRL	48.6	66.4	19.5	24.4	14.7	22.7	22.3	32.8	23.9	28.5
1.0×TRL	49.3	71.8	11.3	14.1	7.6	11.8	14.6	17.5	34.1	43.3
2.0×TRL	37.8	47.7	17.6	20.1	10.7	13.8	18.1	25.4	38.8	49.0
5.0×TRL	28.4	34.9	17.9	23.4	19.3	29.5	11.6	17.7	23.0	29.5
10.0×TRL	22.9	27.4	11.2	12.5	10.5	14.9	12.3	16.1	23.6	30.2

^a Units = percent. The Total Error is defined by Equation (2) in the text. Two values of the factor t in that equation are considered in this report: 2.0 and 3.182, as discussed in Section 2.4. The two columns for each matrix give the corresponding values of the Total Error.

TABLE 17 Total Error Values: GD Data^a

Spiking Level	NCM		NCDF		SCDF		NCS		SCS	
	(2.0)	(3.182)	(2.0)	(3.182)	(2.0)	(3.182)	(2.0)	(3.182)	(2.0)	(3.182)
0.5×TRL	22.8	26.2	24.1	32.5	18.9	23.1	39.6	50.4	19.2	24.2
1.0×TRL	19.6	25.8	24.0	32.6	16.6	22.0	34.1	43.4	16.6	20.2
2.0×TRL	15.6	16.6	21.3	33.5	25.7	36.4	29.6	36.6	19.4	24.4
5.0×TRL	17.7	20.2	10.5	15.3	18.3	24.2	28.1	35.6	17.8	21.5
10.0×TRL	18.1	20.7	11.9	18.8	15.3	21.3	29.2	36.0	16.2	21.4

^a Units = percent. The Total Error is defined by Equation (2) in the text. Two values of the factor t in that equation are considered in this report: 2.0 and 3.182, as discussed in Section 2.4. The two columns for each matrix give the corresponding values of the Total Error.

The results are similar for GD. Using 2.0 as the t-parameter value, 19, 6, and 0 GD Total Error values lie in the excellent, acceptable, and unacceptable ranges, respectively. All the NCS data are in the range from 25% to 50%, comprising five out of the six values. Using 3.182, the corresponding counts are 14, 10, and 1; the lone unacceptable value was just marginally unacceptable at 50.4. As before, the associated DQO is considered to have been achieved to a satisfactory degree.

5.2 GENERAL DATA ANALYSIS

5.2.1 Treatment of Outliers

Because the presence of an outlier can distort the results of a statistical analysis, the identification of outliers is an important step in the data review and analysis process. This section describes the procedure used to address outliers in the analytical method certification data.

The QAPP specifies that the dataset for each agent/matrix combination is to be examined for the presence of outliers, and that each outlier found is to be replaced by a value computed according to the procedure given in Appendix Y of the QAPP. Dixon's test for outliers (Snedecor and Cochran 1989; Beyer 1991) is described in Appendix K of the QAPP and is specified for use in identifying outliers.

Snedecor and Cochran (1989) also provide a replacement procedure and a test for outliers in a two-way classification (Stefansky's test) that is consistent with the specified replacement procedure. Stefansky's test was found to be more useful in this study than Dixon's test, and resulted in significantly fewer outliers. Table 18 summarizes the outliers found and the manner in which they were handled.

The main caveat that must be addressed when treating a given dataset as a two-way classification and applying the statistical methods mentioned above is that these procedures, as with all Analysis-of-Variance (ANOVA) methods, assume an additive model for the factors that cause the observed level-to-level and replicate-to-replicate variations. If the additive model is not approximately valid, these methods for identifying and replacing outliers may give misleading results. A test for non-additivity (Tukey's test) (Snedecor and Cochran 1989) shows that the basic datasets generated in this study definitely do not satisfy the additivity assumption directly. However, it is possible to transform the data in such a way that the additivity assumption is valid (Snedecor and Cochran 1989). The transformation method chosen in this study was to simply divide each found value by the associated target spiking level. Tukey's test shows that the transformed datasets satisfied the additivity assumption, and the outlier test and outlier replacement computation procedures were then valid. In other words, although the data did not directly satisfy the additivity assumption, they did when they were expressed as a fraction or

TABLE 18 Outlier Identification and Replacement

Dataset	Value	Outlier Test		Replacement Value		Prescribed Replacement an Outlier (Dixon)?	Remarks
		Dixon	Stefansky	Prescribed	Correct		
GD-NCM	2.49	Yes	No	2.532	2.718	Yes	Replaced
GD-SCDF	5.88	Yes	No	6.255	6.073	No	Replaced
GD-SCS	2.51	Yes	No	2.519	NC ^a	Yes	Not replaced
	6.18	No	Yes	NC	6.805		Replaced
GB-NCDF	9.26	Yes	No	NC	NC		Run repeated
	45.38	Yes	No	NC	NC		-
Rep 1	9.16	Yes	No	12.473	11.541	Yes	Run repeated
	56.75	No	Yes	NC	51.748		-
Rep 2	9.22	Yes	No	10.663	10.721	No	Replaced
GB-NCS	11.02	Yes	No	9.222	9.602	No	Not replaced
	28.76	No	Yes	NC	22.845		Replaced
GB-SCS	3.13	Yes	No	3.938	2.914	Yes	Replaced
GB-NCM	10.44	Yes	No	10.373	10.364	Yes	Not replaced

^aNC = not computed.

percentage recovered. Once a replacement value for the fraction recovered had been calculated, the true replacement "measured" value was obtained by multiplying the fraction recovered by the spiking level. This approach was used in applying Stefansky's test for outliers and in computing any necessary replacement values. The reason given in the QAPP for replacing outliers rather than deleting them is to preserve a sufficient number of statistical degrees of freedom for the subsequent analyses. With four found values available for each spiking level, the replacement of outliers was not necessary to ensure sufficient data for statistical purposes. In this study, the only reason for replacing outliers was to ensure that four values were always available for all spiking levels so that the IRPQAP software could be run. The statistical computational procedures described in the QAPP are in fact valid in the more general case in which different numbers of values are available for each spiking level.

5.2.2 Tests for and Effects of Nonhomogeneity of Variances

The statistical analysis procedures specified in the QAPP and used in this study to evaluate method performance are all based on the assumption that the variance (the square of the standard deviation) of the replicate found values is the same for all spiking levels. The assumption of constant variance is common in statistical work but, as noted previously, chemical analyses represent one area in which this assumption is often not satisfied. It is common to observe that the standard deviation of measured concentrations in a wide variety of matrices increases systematically with the concentration level. This phenomenon was also observed in this

study, and implies that the statistical results, based in part on the assumption of constant standard deviation, are inaccurate to some extent. This section discusses the methods used to examine the extent of the variance nonhomogeneity and the effects that the nonhomogeneity had on the lack-of-fit test, the zero-intercept test, and the CRL computation. These computations correspond to the first three DQOs listed in Table 4.

Table 19 shows the results of three different methods for examining the nonhomogeneity of variance in the certification datasets. Two methods — Cochran's test (Beyer 1991; Taylor 1987) and Levene's test (Snedecor and Cochran 1989) — evaluate sets of variance values for consistency without regard to their order. The third method was to carry out a linear regression of standard deviation against spiking level and to examine the statistical significance of the resulting regression line. As the table shows, these tests gave reasonably, but not totally, consistent results.

Of the three tests, the regression line check is probably the most sensitive to the expected behavior, and gave a statistically significant positive result in all but two cases. In several cases, the regression was highly linear, with R^2 values in excess of 0.95. For some reason, GD gave more consistently linear regressions than GB, but even for the two datasets for which the linear regression is not statistically significant (GB-NCDF and GB-SCDF), the standard deviation was observed to increase erratically with spiking level. The results from the other tests are reasonably

TABLE 19 Tests for Nonhomogeneity of Variances^a

Dataset	Linear Dependence of S on Target			Levene's Test			Cochran's Test	
	t (3df)	Significant?	R^2	F	df	Significant?	Test Value	Significant?
GD-NCM	7.465	Highly	0.949	4.58	4, 14	Yes	0.7629	Highly
GD-NCS	36.026	Extremely	0.998	6.78	4, 15	Highly	0.7310	Highly
GD-SCDF	3.637	Yes	0.815	3.28	4, 14	Yes	0.8525	Highly
GD-SCS	16.331	Extremely	0.989	47.54	4, 14	Extremely	0.8288	Highly
GD-NCDF	6.032	Highly	0.924	6.12	4, 15	Highly	0.7931	Highly
GB-SCS	12.308	Highly	0.981	2.86	4, 14	No	0.7316	Highly
GB-NCS	7.936	Highly	0.955	1.69	4, 14	No	0.6467	Yes
GB-NCDF	1.245	No	0.341	4.36	4, 14	Yes	0.7712	Highly
GB-SCDF	2.581	No	0.690	3.84	4, 15	Yes	0.5695	No
GB-NCM	4.989	Yes	0.892	4.26	4, 15	Yes	0.4991	No

^a Terminology used for describing statistical significance:

No: 95% confidence value > test statistic

Yes: 99% value > test statistic $\geq$ 95% value

Highly: 99.9% value > test statistic $\geq$ 99% value

Extremely: test statistic > 99.9% value (99.9% confidence value not available for Cochran's test).

consistent for GD, but not for GB. Both tests indicate that the variances for GD are not homogeneous, with degrees of confidence varying from 95% to beyond 99.9%. However, for GB, the other two tests agreed only for one dataset (GB-NCDF); Levene's test indicated that the data do not support the conclusion of nonhomogeneity for GB-SCS and GB-NCS, in contrast to the results from Cochran's test. For GB-SCDF and GB-NCM, the results were just the reverse, with Cochran's test indicating homogeneity. Note that for all datasets, at least one of the three tests indicated nonhomogeneity in the variances.

The conclusion drawn from the results presented in Table 19 is that the method standard deviations (the square roots of the variances) for the methods described in this report show a systematic increase with agent concentration. The assumption of constant standard deviation or, equivalently, constant variance is not valid for these data.

The effect of nonhomogeneity of variances on the results from the computations used for method evaluation (carried out by using IRPQAP in accordance with the descriptions in the QAPP) depends on the specific calculation. The lack-of-fit test is relatively unaffected because both the numerator and the denominator used in the F-test statistic are insensitive to any systematic trends in variance. The zero-intercept test would be affected more strongly. That test, as described in the QAPP, is equivalent to a test that computes a 95% confidence region for the intercept and checks to see if the origin is included. The standard deviation is used in the computation. If the measurement standard deviation increases significantly with spiking level, the appropriate standard deviation to use is either that associated with the lowest spiking level or an extrapolated zero-concentration value. In either case, the appropriate standard deviation value may be substantially less than the overall average value, especially if the dependence on spiking level is strong, and the corresponding zero-intercept test would be more stringent than one in which the standard deviation is assumed to be constant.

The other main result obtained from IRPQAP is the value of the CRL. Equation (4) shows that the CRL is essentially proportional to the standard deviation of found values about the regression line. When the method standard deviation increases significantly with spiking level, the greatest contribution to $s_{y,x}$ comes from the higher spiking levels rather than from results at or below the TRL. The CRL that would be obtained if the variation of standard deviation with spiking level is taken into account would be smaller than if this variation is not accounted for.

An example of the difference that a more realistic treatment of the standard deviation might make was generated for the GD-NCS dataset. This dataset showed the strongest linear relationship between standard deviation and spiking level of all GB or GD datasets. Application of more general formulas derived without the assumption of constant standard deviation resulted in confidence bands about regression that described the data much more accurately and yielded a CRL value of 1.7 ng/g instead of the 4.6 ng/g reported in Table 13.

The conclusion drawn from these considerations is that different assumptions regarding the statistical behavior of the experimental data, particularly the method standard deviation, can significantly affect the results of the zero-intercept test and the CRL computation.

5.2.3 GB and GD Percent Recovery Data

An unexpectedly large number of certification measurements for GB and GD had recovery percentages well in excess of 100%. It is not clear whether this finding indicates a true positive bias in these methods or whether there was some unknown problem with the technique or the instrumentation. Several problem areas were identified. The sample spiking and extraction procedures can be eliminated as the source of the problem because the solvent spike results, which involve no metal, soil, or decon fluid spiking and extraction steps, also provided systematically high percent recoveries. Other possibilities include the following: the procedures used to prepare the agent stock and working stock solutions for both the calibration and matrix spiking solutions; possible memory effects caused by deposition of agent on various surfaces in the analytical train (ATD focusing tube, transfer line, and column, for example) and subsequently release over time (depending on usage); variability in detector response caused by daily operation or changes in the laboratory environment (e.g., temperature, humidity, line voltage); and degradation of the agent caused by reaction with substances such as water vapor or nitrogen dioxide in the ambient air used in the transfer of agent from the extract to a sorbent cartridge. In support of the latter possibility, the laboratory temperature and relative humidity changed noticeably during the course of the day during the period from May to October when this work was carried out; we found that ambient humidity had a significant negative effect on the recovery of HD in experimental work conducted after the work described in this report. Because the experimental results were of sufficient quality to easily certify the methods according to the procedures specified in the USATHAMA QAPP, these hypotheses were not investigated experimentally.

In an attempt to narrow down the possible causes using available data, the certification and associated QC data were reviewed in some detail. Analysts examined the behavior of the data over time and compared the QC sample results with the certification sample results. An ANOVA was carried out on the certification sample results in an attempt to isolate possible sources of variation. The results of this examination are described in Appendix C.

5.3 SUMMARY OF QUALITY CONTROL SAMPLE RESULTS

The number and type of QC samples used during the method certification runs were described in Section 3.2.2 and summarized in Table 6, which also presents the QC acceptance criteria that were adopted. The analyses results for those samples are summarized in this section.

Two runs were repeated for GB-NCDF because of problems with outliers, and one run for GD-NCS was repeated because of problems with the QC sample results. In the latter case, the problems were apparently caused by contamination of the sorbent cartridges; both matrix blank samples as well as a separate blank sorbent cartridge all gave discernible GD peaks of about 0.2 ng, and the low calibration QC sample result was also too high. No other certification runs for either GB or GD needed to be repeated.

Table 20 summarizes the QC sample analysis results for the runs used in the certification data analyses. The numerical values represent average percent differences for matrix and solvent spike duplicates and average percent deviations for the calibration and check standards. The numbers in parentheses represent the total number of samples of the given type that were analyzed.

A small GB peak (calibrated value = - 0.04 ng) was seen in one of the forty GB matrix blank samples analyzed, and a small GD peak (calibrated value = 0.2 ng) was seen in one of the forty GD matrix blank samples. One GB low calibration check and two GB solvent spike duplicates were outside of the adopted QC acceptance criteria. Three GD solvent spike duplicates were also outside of the adopted criteria. All other QC results were within the adopted criteria (described in Table 6).

TABLE 20 Summary of Method Certification Quality Control Sample Results^a

QC Sample Name	Spiking Level	GB	GD
Matrix blank (two used per run)	0×TRL	A discernible peak seen in 1 out of 40 samples ^b	A discernible peak seen in 1 out of 40 samples ^c
Matrix spike duplicates	10×TRL	Metal substrate: 1.96 (4) Decon fluids: 2.79 (8) Soils: 6.05 (8)	Metal substrate: 2.18 (4) Decon fluids: 2.73 (8) Soils: 3.36 (8)
Solvent spike duplicates	1×TRL	6.53 (20)	7.85 (20)
Low calibration standard	0.375×TRL	7.18 (20)	8.40 (20)
High calibration standard	12.5×TRL	6.03 (20)	5.38 (20)
Check standard	8×TRL	6.54 (19)	5.72 (19)

^a Numerical values represent average percent differences for matrix and solvent spike duplicates and average percent deviations for the calibration and check standards. Numbers in parentheses represent the total number of samples analyzed.

^b Calibrated amount = - 0.04 ng.

^c Calibrated amount = 0.2 ng.

6 SUMMARY AND CONCLUSIONS

In response to the U.S. Army's need for highly sensitive analytical methods for agents GB and GD in selected waste matrices, ANL has developed methods for the analysis of metal substrate, spent hypochlorite decon fluid, and soil matrices for these agents. These methods utilize sorbent cartridge preconcentration and thermal desorption techniques combined with gas chromatography using flame photometric detection to achieve the desired sensitivity and specificity.

The methods were validated using test procedures adopted from the USATHAMA Quality Assurance Program Plan (USATHAMA 1990) to demonstrate that the analytical processes could be adequately calibrated (method precertification phase) and could achieve the desired level of sensitivity and reproducibility over a minimum of four different days (method certification phase). The methods described in this report met all the requirements set forth in the USATHAMA QAPP for demonstrating method validity when applied to the test matrices (metal substrate, spent hypochlorite decon fluid, and soil). On the basis of the test results, the methods described in this report are capable of quantitatively determining levels of GB and GD in the low ppb range in all sample matrices tested.

The CRLs achieved by these methods range from 8.3 to 19 ppb for GB and from 1.8 to 5.3 ppb for GD. All the CRLs are less than the corresponding TRLs. MDLs range from 1.7 to 8.2 ppb for GB and from 0.39 to 1.2 ppb for GD. The sorbent cartridge preconcentration and thermal desorption technique used is believed to be capable of providing even greater sensitivity, if necessary, by increasing the total volume of sample extract that is used. The volume of extract can be increased either by injecting a larger volume, up to a maximum of 150 mL, onto the fuzz tube/sorbent cartridge apparatus, or by using multiple injections. The methods also provide adequate protection against both false positive and false negative errors in determining the level of agent with respect to an assumed action level. Finally, the methods are acceptable with respect to the Total Error criterion originally proposed by McFarren et al. (1970).

The regression of found versus target agent concentrations was linear for each agent-matrix combination (the minimum value of R^2 found was 0.9954), and all regression lines passed well within 95% confidence limits about the origin. The standard deviation of measured values about the regression line was found to increase with target concentration for all agent/matrix combinations; in all but two cases, a linear regression of the standard deviation against target spiking level was significant at the 95% confidence level or higher. The observed linear increase of standard deviation with spiking level is common in analytical chemistry.

The slopes of the regression lines of found versus target concentrations were statistically greater than unity at the 95% confidence level in all but one case, averaging 1.089 over all agent/matrix combinations. The reason for this apparent positive bias is unknown, although the fact that the recovery percentages from the solvent spike QC samples were also systematically greater than unity suggests that it does not lie within the spiking or extraction steps. Statistically detectable day-to-day fluctuations occurred in the recovery percentages, although they were not

of sufficient magnitude to cause a method to fail to achieve the data quality objectives and do not explain the systematic positive bias. These fluctuations are attributed to day-to-day variations in the instrument response.

The treatment of outliers recommended in the USATHAMA QAPP was modified in this study. A consistent test and replacement procedure was identified and used. Spreadsheets were designed and used to implement the appropriate statistical procedures in cases involving outliers.

7 RECOMMENDATIONS

The analytical methods developed by ANL for GB and GD in selected waste matrices have been shown to be valid for their intended applications, as described in this report. On the basis of our findings, we have developed several recommendations for further work in this area, as described in this section.

As indicated in Section 2.1, because it is impossible to anticipate all potential technical problems that might arise when an analytical method is used to analyze real-world samples, method validation is always an incomplete and ongoing process. As shown in Table 1, a wide variety of substrates in addition to unpainted metal are used in the testing at DPG. Also, several different types of decon solution are in use at DPG and at other U.S. Army facilities. Finally, soil types vary widely in their physical and chemical properties, and these differences may affect the results of analyses conducted by using the methods described in this report. We recommend, then, that these methods be further tested on a wider variety of sample matrices, including (if possible) real-world samples known to be contaminated with GB or GD or their degradation products.

It is widely understood in the analytical chemistry community that for an analytical method to become a standard accepted by the community, it should be tested by more than one laboratory. This type of additional testing is recognized as part of the overall method validation process. We therefore recommend that the methods developed by ANL for this study be tested further by one or more independent laboratories.

The analytical methods described in this report make use of sorbent cartridge preconcentration and thermal desorption techniques for delivering a sample to the GC in order to achieve significantly greater sensitivity than that attainable by using direct liquid injection methods. Since this work was begun, a new system for sample introduction, called a high-volume liquid injection (HVL) system, has become commercially available. This system also promises significant gains in sensitivity, comparable to those achieved by the methods used in this study. The sorbent cartridge preconcentration and thermal desorption techniques used here suffer from the disadvantages that they are labor-intensive and technique-sensitive. The manipulations must be carried out manually, require the full attention of the laboratory analyst, and are subject to variations in the laboratory environment (temperature, humidity, etc.). The HVL system is automated, requires much less direct manipulation from laboratory personnel, and eliminates one of the two thermal desorption steps. Because of these features, greater reproducibility and consistency might be expected using this system rather than the methods used in this study. Greater consistency would result directly in lower CRLs for the method, all other things being equal. In addition, it may be possible to inject larger volumes of sample extract into the HVL, thereby increasing the sensitivity of the method. We therefore recommend that the methods described in this report be modified to incorporate the HVL technology and that the modified method be tested on the sample matrices used in our study.

REFERENCES

ANL: see Argonne National Laboratory

Argonne National Laboratory, 1995, *Toxics Testing Performance Evaluation Plan*, prepared for U.S. Army Test and Evaluation Command (Nov. 29).

Beyer, W.H. (Ed.), 1991, *CRC Standard Probability and Statistics Tables and Formulae*, CRC Press, Boca Raton, Fla.

EPA: see U.S. Environmental Protection Agency

Kayser, R., 1988, EPA Office of Solid Waste and Emergency Response, Variances Section, Memorandum to Mr. Terry Grogan, Chief, Variances Section, Re: *Verification Data Submitted by Syntex* (Oct. 14).

McFarren, E.F., R.J. Lishka, and J.H. Parker, 1970, "Criterion for Judging Acceptability of Analytical Methods," *Analytical Chemistry* 42, 358-365.

Mosteller, F., and J.W. Tukey, 1977, *Data Analysis and Regression*, Addison-Wesley Publishing Company, Reading, Mass.

Potomac Research, Inc., 1990, *User's Manual: The Installation Restoration Project Quality Assurance Programs (Pre-Certification and Certification) for the IBM PC/XT or PC/AT*, prepared for the U.S. Army Toxic and Hazardous Materials Agency, Aberdeen Proving Ground, Md. (revised Aug. 31).

Snedecor, G.W., and W.G. Cochran, 1989, *Statistical Methods*, Eighth Edition, Iowa State University Press, Ames, Iowa.

Southern Research Institute, 1992, *3X/5X Delisting Plan — Analytical Methods Report*, Final Report, Phase I (March 10).

Southern Research Institute, 1992, *3X/5X Waste Sampling Test Report*, Final Report (July 31).

Southern Research Institute, undated, *Analytical Methods Test Plan for the 3X/5X Delisting Plan*.

SRI: see Southern Research Institute

Taylor, J.K., 1987, *Quality Assurance of Chemical Measurements*, Lewis Publishers, Inc., Chelsea, Mich.

U.S. Army Toxic and Hazardous Materials Agency (now the Army Environmental Center), 1990, *Quality Assurance Program*, U.S. Army Toxic and Hazardous Materials Agency, Aberdeen Proving Ground, Md., Report No. USATHAMA PAM 11-41 (Jan.).

USATHAMA: see U.S. Army Toxic and Hazardous Materials Agency

U.S. Environmental Protection Agency, 1986, *Test Methods for Evaluating Solid Waste*, Office of Solid Waste and Emergency Response, Report No. SW-846 (Nov.).

U.S. Environmental Protection Agency, 1995, *Guidance for Methods Development and Methods Validation for the RCRA Program*, Office of Solid Waste and Emergency Response (June 16).

APPENDIX A: GLOSSARY OF TERMS

Action Level — concentration of a chemical agent in wastes and residues below which these wastes do not pose a significant hazard.

Automated Thermal Desorption — an automated technique for introducing a sample into a gas chromatograph by thermally desorbing it from a sorbent resin.

Average Found Concentration — the average concentration found from a large number (in principle, an infinite number) of analyses on the same sample.

Bias — the analytical method bias; $B = AFC - TC$.

Certified Reporting Limit (method) — the true concentration at which there is less than a 5% chance of obtaining a found concentration below the value that can just be distinguished from zero at the 95% confidence level (based on USATHAMA 1990).

Data Quality Objective — qualitative or quantitative objectives of precision, accuracy, and completeness.

Filtered, Reduced, pH-Adjusted, and Diluted — sample preparation conducted for decontamination fluids prior to chemical analysis.

High-Test Hypochlorite — a commercially available granular mixture containing 65% calcium hypochlorite and 35% inert ingredients.

Installation Restoration Program — program for investigation and cleanup of contaminated sites at military installations.

Installation Restoration Program Quality Assurance Program (IRPQAP) — software used by the U.S. Army Environmental Center for method validation and other purposes.

Method Detection Limit — the minimum concentration of a substance that can be measured and reported with 99% confidence that the analyte concentration is greater than zero; this limit is determined from analysis of a sample in a given matrix type containing the analyte (USEPA 1986).

Never-Contaminated — a sample matrix that is known never to have been previously contaminated with agent.

R^2 — a measure of the quality of the fit between measurement results and an assumed mathematical model (e.g., a linear or quadratic curve). R^2 is less than or equal to one, with $R^2 = 1$ indicating a perfect fit.

S — an estimate of the method standard deviation at a particular concentration or spiking level.

Surrogate — a substance that is chemically stable, but similar in structure and behavior to a target analyte; used to provide a quality control check for unusual matrix effects and gross errors in the solvent extraction step of the analytical process.

Suspect-Contaminated — a soil matrix that may have been previously exposed to agent, or to a decontamination fluid matrix that is known to have been previously exposed to agent.

Target Reporting Level — the desired, required, or expected method reporting limit estimated before performing certification analyses to determine target concentrations during certification (USATHAMA 1990).

Tolerance Level — a concentration decision level less than the Action Level that is chosen to provide the necessary safeguard against false negative decisions.

True Concentration — the actual concentration of agent in a sample.

APPENDIX B: SUMMARY OF CERTIFICATION DATA

Tables B.1 through B.10 present summaries of the certification data obtained for GB and GD for the sample matrices used in this study. Each table gives the spiking level, expressed as a multiple of the TRL; the associated amount of agent spiked into the sample, in nanograms (ng); and the amount of agent recovered, in ng, for each spiking level and for each of the four runs used in the method certification statistical analyses. The tables also list the dates on which the four runs were carried out. Those values that represent outlier replacement values are flagged and the original outlier value is given in a footnote to each table, as needed.

TABLE B.1 Certification Data for GB on Previously Noncontaminated Metal (NCM)

Spiking Level	Amount Spiked (ng)	Amount Recovered (ng)			
		10/1/96	10/2/96	10/2/96	10/7/96
0.5×TRL	2.40	2.78	2.40	2.90	3.28
1.0×TRL	4.80	5.44	4.20	5.29	6.43
2.0×TRL	9.60	12.17	10.44	11.84	12.05
5.0×TRL	24.00	28.77	26.94	27.36	29.78
10.0×TRL	48.00	56.52	53.29	57.23	54.48

TABLE B.2 Certification Data for GB in Previously Noncontaminated Decon Fluid (NCDF)

Spiking Level	Amount Spiked (ng)	Amount Recovered (ng)			
		5/23/96	5/29/96	5/30/96	8/22/96
0.5×TRL	2.40	2.66	2.63	2.81	2.58
1.0×TRL	4.80	4.97	5.12	5.24	5.15
2.0×TRL	9.60	11.06	11.07	10.71	10.72133 ^a
5.0×TRL	24.00	27.55	25.51	26.17	24.96
10.0×TRL	48.00	52.84	52.04	51.59	52.55

^a Outlier replacement value; outlier value = 9.22 ng.

TABLE B.3 Certification Data for GB in Suspect-Contaminated Decon Fluid (SCDF)

Spiking Level	Amount Spiked (ng)	Amount Recovered (ng)			
		7/17/96	7/19/96	7/22/96	7/23/96
0.5×TRL	2.40	2.38	2.18	2.58	2.36
1.0×TRL	4.80	4.85	4.62	5.03	4.82
2.0×TRL	9.60	10.37	9.90	9.91	10.31
5.0×TRL	24.00	22.39	24.12	27.35	24.12
10.0×TRL	48.00	48.81	49.94	51.73	47.57

TABLE B.4 Certification Data for GB in Previously Noncontaminated Soil (NCS)

Spiking Level	Amount Spiked (ng)	Amount Recovered (ng)			
		8/1/96	8/5/96	8/6/96	8/7/96
0.5×TRL	2.40	2.20	2.63	2.53	2.67
1.0×TRL	4.80	5.13	5.31	5.19	5.40
2.0×TRL	9.60	9.68	9.96	11.02	9.98
5.0×TRL	24.00	22.84458 ^a	24.82	23.84	25.71
10.0×TRL	48.00	49.01	50.77	50.89	52.76

^a Outlier replacement value; outlier value = 28.76 ng.

TABLE B.5 Certification Data for GB in Suspect-Contaminated Soil (SCS)

Spiking Level	Amount Spiked (ng)	Amount Recovered (ng)			
		9/30/96	10/1/96	10/2/96	10/3/96
0.5×TRL	2.40	2.78	2.914167 ^a	2.69	2.77
1.0×TRL	4.80	6.04	5.94	5.23	5.52
2.0×TRL	9.60	11.61	12.83	10.90	11.34
5.0×TRL	24.00	28.16	26.38	25.22	27.69
10.0×TRL	48.00	54.83	56.67	50.35	54.27

^a Outlier replacement value; outlier value = 3.13 ng.

TABLE B.6 Certification Data for GD on Previously Noncontaminated Metal (NCM)

Spiking Level	Amount Spiked (ng)	Amount Recovered (ng)			
		7/17/96	7/18/96	7/22/96	7/23/96
0.5×TRL	0.60	0.72	0.71	0.68	0.70
1.0×TRL	1.20	1.33	1.23	1.38	1.30
2.0×TRL	2.40	2.76	2.71	2.73	2.71817 ^a
5.0×TRL	6.00	6.87	6.65	6.94	6.80
10.0×TRL	12.00	13.33	13.81	13.91	13.62

^a Outlier replacement value; outlier value = 2.49 ng.

TABLE B.7 Certification Data for GD in Previously Noncontaminated Decon Fluid (NCDF)

Spiking Level	Amount Spiked (ng)	Amount Recovered (ng)			
		5/23/96	5/29/96	5/29/96	5/30/96
0.5×TRL	0.60	0.72	0.65	0.62	0.65
1.0×TRL	1.20	1.29	1.24	1.28	1.44
2.0×TRL	2.40	2.02	2.57	2.50	2.44
5.0×TRL	6.00	5.83	6.39	6.09	6.27
10.0×TRL	12.00	11.78	12.32	12.80	11.16

TABLE B.8 Certification Data for GD in Suspect-Contaminated Decon Fluid (SCDF)

Spiking Level	Amount Spiked (ng)	Amount Recovered (ng)			
		8/6/96	8/7/96	8/8/96	8/12/96
0.5×TRL	0.60	0.64	0.69	0.67	0.68
1.0×TRL	1.20	1.31	1.28	1.35	1.22
2.0×TRL	2.40	2.50	2.79	2.73	2.32
5.0×TRL	6.00	6.59	6.77	6.57	6.07250 ^a
10.0×TRL	12.00	12.82	12.62	13.28	11.84

^a Outlier replacement value; outlier value = 5.88 ng.

TABLE B.9 Certification Data for GD in Previously Noncontaminated Soil (NCS)

Spiking Level	Amount Spiked (ng)	Amount Recovered (ng)			
		7/31/96	8/1/96	8/6/96	9/19/96
0.5×TRL	0.60	0.80	0.74	0.68	0.69
1.0×TRL	1.20	1.55	1.43	1.35	1.35
2.0×TRL	2.40	2.99	2.90	2.69	2.73
5.0×TRL	6.00	7.37	7.05	6.47	6.85
10.0×TRL	12.00	14.95	14.39	13.38	13.79

TABLE B.10 Certification Data for GD in Suspect-Contaminated Soil (SCS)

Spiking Level	Amount Spiked (ng)	Amount Recovered (ng)			
		9/18/96	9/18/96	9/19/96	9/20/96
0.5×TRL	0.60	0.63	0.67	0.67	0.69
1.0×TRL	1.20	1.29	1.30	1.34	1.37
2.0×TRL	2.40	2.51	2.72	2.71	2.71
5.0×TRL	6.00	6.45	6.64	6.8050 ^a	6.87
10.0×TRL	12.00	12.47	12.42	13.32	13.38

^a Outlier replacement value; outlier value = 6.18 ng.

APPENDIX C: GB AND GD PERCENT RECOVERY INVESTIGATION

As indicated in Section 5.2.3, an attempt was made to identify the causes of the apparent upward bias exhibited by the methods described in this report. This investigation consisted of various statistical analyses performed on the certification datasets and is described in this appendix.

The investigation was conducted using the values of the ratio (r) of the found concentration to the target spiking level; r equals the percent recovery divided by 100. This procedure was necessary for two reasons. First, in nearly all cases, the standard deviations of the measurements clearly depended on the spiking level, and most standard statistical procedures assume that the standard deviation is constant (i.e., the data are homoscedastic). Dividing the analytical results by the spiking level made the data much more homoscedastic. Secondly, as noted in Section 5.2.2, standard ANOVA methods assume that the effects being sought are additive. We found that the basic certification data did not satisfy this additivity assumption, but that data scaled by dividing by the spiking level did satisfy the assumption.

BEHAVIOR OVER TIME

Figures C.1 and C.2 present the certification analysis data for GB and GD over the period from May 1996 to October 1996. The mean value of r for the certification analyses for each day's run is shown using a filled diamond symbol, and the high and low values are represented by a vertical bar with horizontal caps on each end. The GB data are relatively consistent until the September and October runs, when both the day-to-day variability and the mean recovery percentages increase noticeably. The daily GD data reveal a fairly systematic decrease in variability over time, but considerable day-to-day variability is apparent. The mean GD recovery percentage is approximately constant over time, except for a substantial upward deviation at about day 210. No long-term trends are apparent in either dataset. It is also significant that no systematic or periodic short-term trends are apparent. If the high percent recoveries were caused by evaporative concentrations of stock agent solutions, a sawtooth-like periodicity might be apparent in the data. However, no such effect can be seen.

Figures C.3 and C.4 show the behavior of the GB and GD QC sample recovery percentages over time and, for comparison, also show the daily mean certification recovery percentages. It is apparent that the GB QC sample data consistently segregate themselves to a significant degree. The solvent spike recovery percentages are generally higher than those from the certification samples, and the calibration checks (including the check standard sample) tend to be lower than the certification data. This result is confirmed by statistical significance tests described in the next section. The tendency for segregation is not as apparent in the GD data, although it seems to be present in the data collected between days 195 (July 13) and 225 (August 12) but not in the other two clusters of data shown.

FIGURE C.1 GB Certification Data Percent Recovery versus Date

FIGURE C.2 GD Certification Data Percent Recovery versus Date

FIGURE C.3 GB Quality Control Data Percent Recovery versus Date

FIGURE C.4 GD Quality Control Data Percent Recovery versus Date

QC DATA COMPARISONS

Tables C.1 and C.2 present statistical summaries of the GB and GD percent recovery data. The data have been separated into the following categories: solvent spike data, check standard data, high calibration data, low calibration data, and certification data. The first four categories contain results from the indicated QC samples.

Several comparisons between the GB and GD data were made using ANOVA methods (Snedecor and Cochran 1989). Prior to comparing mean recoveries for the different data categories, the variances for each agent in each data category were compared with each other using standard F tests. For this purpose, data from the three calibration QC check samples (high value, low value, and check standard) were pooled.

For the GD data, the variance in the solvent spike data is statistically higher than the variance in the calibration QC check data, but the variance in the certification data is not

TABLE C.1 Summary of GB Percent Recovery Data

Parameter	Solvent Spikes	Check Standards	High Calibration Samples	Low Calibration Samples	Certification Samples
Number of samples	47	29	28	28	100
Mean	117.71	97.77	94.72	95.18	110.13
Standard deviation	8.75	10.08	9.58	11.40	8.83
Maximum value	143.33	120.76	112.90	114.44	136.70
Minimum value	100.21	70.96	74.32	68.89	87.50
Standard deviation of the mean	1.28	1.87	1.81	2.15	0.88

TABLE C.2 Summary of GD Percent Recovery Data

Parameter	Solvent Spikes	Check Standards	High Calibration Samples	Low Calibration Samples	Certification Samples
Number of samples	40	20	21	20	100
Mean	111.23	110.13	104.97	108.22	110.72
Standard deviation	9.29	3.63	6.43	6.88	7.44
Maximum value	130.00	114.69	119.33	124.44	133.33
Minimum value	92.50	102.19	94.27	95.56	84.17
Standard deviation of the mean	1.47	0.81	1.40	1.54	0.74

statistically different from the variances in either the solvent spike or the calibration check data. None of the GB variances are significantly different. These results indicate that, with the exception of the GD solvent spike data, the variances for each agent may be pooled to estimate the standard error of a given combination of means.

After examining the variances, the following comparisons were made for each agent: (1) the solvent spike data with the mean of the three calibration check categories, (2) the certification data with the mean of the three calibration check categories, and (3) the certification data with the solvent spike data. For each comparison, we computed the value of the indicated difference of mean values and compared it with the product of the appropriate Student's-t value and the estimated standard error of the difference. If the difference in means is larger than t times the standard error, the difference is significant at the confidence level associated with the value of t that was used. The estimated standard errors were based on the appropriate pooled estimates of the corresponding variances.

Tables C.3 and C.4 summarize the results of these computations. The results show that all the tested differences for GB are extremely significant (99.9% confidence or higher). For GD, the calibration check mean was significantly lower (> 99.9% confidence) than either the solvent spike or the certification mean. No significant difference between the GD certification and solvent spike means was apparent.

The means for the three calibration checks were also compared using a Least Significant Difference approach (Snedecor and Cochran 1989). For GB, no statistically significant difference was observed for any pair of calibration check mean values. For GD, the check standard percent recovery was statistically higher (95% confidence) than the high calibration check percent recovery, which in turn was statistically higher than the low calibration check percent recovery.

The statistical significance of these comparisons is sufficiently high that they undoubtedly represent real differences. One might expect the solvent spike data to be higher than the other

TABLE C.3 Comparisons of GB Percent Recovery Data

Quantity 1	Quantity 2	Comparison Value ^a	t Times Standard Error	Confidence Level (%)
Solvent spike mean	Mean of the three calibration checks	21.82	3.38	99.9
Certification mean	Mean of the three calibration checks	14.24	2.75	99.9
Solvent spike mean	Certification mean	7.58	5.23	99.9

^a Comparison value = Quantity 1 – Quantity 2.

TABLE C.4 Comparisons of GD Percent Recovery Data

Quantity 1	Quantity 2	Comparison Value ^a	t Times Standard Error	Confidence Level (%)
Solvent spike mean	Mean of the three calibration checks	3.46	2.95	99.9
Certification mean	Mean of the three calibration checks	2.94	2.13	99.9
Solvent spike mean	Certification mean	0.51	2.96	Not significant

^a Comparison value = Quantity 1 – Quantity 2.

data because the solvent spike involves injection of agent directly into the solvent used for extractions and does not involve spiking an agent into some matrix from which it is then extracted. The extraction efficiency in all cases is expected to be less than 100%, and the solvent spike data are therefore expected to be systematically higher than the rest. The explanations for the other differences are not as obvious. The fact that the solvent spike recovery percentages are high and suffer from the same day-to-day variability as the rest of the data suggests that the reason for recovery percentages in excess of 100% does not lie in the spiking and extraction processes. The fact that the GB and GD recovery percentages differ somewhat is probably associated with the use of different concentrations for the two agents in corresponding samples.

ANOVA OF CERTIFICATION PERCENT RECOVERIES

ANL completed a two-way (random effects model) analysis of variance (Snedecor and Cochran 1989) on the recovery percentages for the certification sample data for GB and GD. The analysis was conducted to (1) evaluate the statistical significance of the day-to-day variability observed in the data, and (2) assess the significance of any possible differences between different spiking levels. Table C.5 shows the results of this investigation for GB and GD. Data from eleven agent/matrix combinations are shown (the GB-NCM run was repeated, and both the original and the repeat data are shown). The following information is shown for each agent/matrix combination: the daily mean percent recovery (certification analysis samples only), the overall mean for the dataset, the F-ratios and degrees of freedom associated with the four replicates and the five different spiking levels, the significance of the F-ratios, and the slope of the regression line of found concentrations against spiking levels. The F-ratios are measures of the statistical significance of the day-to-day and level-to-level differences in percent recovery.

The replicate F-ratios are rather variable. Of the eleven datasets, the day-to-day variability (as indicated by the replicate F-ratio) is highly or extremely significant in six, significant in one,

TABLE C.5 GB and GD ANOVA Results

Dataset	Percent Recovery					Replicate F-Ratio ^a (df)	Significant?	Regression Slope	Levels	
	Day 1	Day 2	Day 3	Day 4	Mean				F-Ratio ^a (df)	Significant?
GD-NCDF	1.014	1.056	1.045	1.055	1.042	0.32 (3,12)	No	1.0001	1.73 (4,12)	No
GD-NCM	1.143	1.119	1.147	1.111	1.135 ^b	0.87 (3,11)	No	1.1394	2.56 (4,11)	No
GD-NCS	1.141	1.269	1.202	1.115	1.181	58.43 (3,12)	Extremely	1.1737	4.08 (4,12)	Yes
GD-SCDF	1.135	1.106	1.074	1.003	1.081 ^b	26.42 (3,11)	Extremely	1.0548	3.55 (4,11)	Yes
GD-SCS	1.057	1.095	1.121 ^b	1.136	1.102	17.66 (3,11)	Extremely	1.0776	1.00 (4,11)	No
GB-NCDF	1.109	1.080 ^b	1.093	1.109	1.098	0.98 (3,11)	No	1.0871	2.76 (4,11)	No
GB-NCM	1.200	1.198	1.250	1.125	1.193	7.66 (3,12)	Highly	1.1780	2.54 (4,12)	No
GB-NCM rep	1.187	1.039	1.175	1.267	1.167	7.05 (3,12)	Highly	1.1556	0.85 (4,12)	No
GB-SCDF	1.007	0.990	1.075	1.012	1.021	3.28 (3,12)	No	1.0309	1.15 (4,12)	No
GB-NCS	0.993 ^b	1.066	1.067	1.090	1.054	4.89 (3,11)	Yes	1.0527	1.97 (4,11)	No
GB-SCS	1.188	1.214 ^b	1.089	1.154	1.161	8.74 (3,11)	Highly	1.1198	3.91 (4,11)	Yes

^a Critical F-Ratio Values:

3 and 11 df: (95%) 3.59; (99%) 6.22; (99.9%) 11.56

3 and 12 df: (95%) 3.49; (99%) 5.95; (99.9%) 10.80

4 and 11 df: (95%) 3.36; (99%) 5.67; (99.9%) 10.35

4 and 12 df: (95%) 3.26; (99%) 5.41; (99.9%) 9.63

^b One outlier replaced.

and not significant in four. In the table, an F-ratio is designated extremely significant if it is larger than the 99.9% confidence critical value, highly significant if it is greater than the 99% confidence critical value, and significant if it is greater than the 95% confidence critical value. An F-ratio is considered not significant if it is less than the 95% confidence critical value. The critical values for the appropriate number of degrees of freedom are shown in a footnote to the table. The GD-NCS dataset, which has an extremely significant replicate F-ratio of 58.43, does not quite satisfy the ANOVA additivity assumption according to Tukey's test (Snedecor and Cochran 1989), and the apparent extreme value of the F-ratio may be too high for this reason. The results indicate that some factor that fluctuates on a day-to-day basis in a rather intermittent manner significantly affects the analyses. It seems probable that this effect is caused by random variations in the detector sensitivity. A rough correlation exists between the regression slope and the replicate F-ratio; the three highest slopes are all associated with highly or extremely significant fluctuations in the recovery percentage, and the two lowest slopes occur with datasets having no significant variations in percent recovery. The regression slope is highly correlated with the overall mean percent recovery, as it should be.

The F-ratios for the spiking levels are generally not significant. The three datasets in which significant F-ratios were found for the spiking levels are also three of the datasets having the highest four replicate F-ratios, and this association may indicate that the significance of these levels F-ratios may be somewhat overstated. The non-significance of the spiking levels effects implies that the recovery percentages do not depend significantly on the spiking level.

M98004345

Report Number (14) ANL/EAD/TM--76

Publ. Date (11) 197710

Sponsor Code (18) DOD, XF

UC Category (19) UC-~~000~~, DOE/ER

DOE