United States Environmental Protection Agency
Science Advisory Board Washington, DC
EPA-SAB-RAC-97-008 September, 1997

AN SAB REPORT: REVIEW OF THE MULTI-AGENCY RADIATION SURVEY AND SITE INVESTIGATION MANUAL (MARSSIM)

PREPARED BY THE RADIATION ADVISORY COMMITTEE (RAC) OF THE SCIENCE ADVISORY BOARD

September 30, 1997

EPA-SAB-RAC-97-008

Honorable Carol M. Browner
Administrator
U.S. Environmental Protection Agency
401 M Street, S.W.
Washington, DC 20460

Re: Review of the Multi-Agency Radiation Survey and Site Investigation Manual (MARSSIM) draft dated December 1996

Dear Ms. Browner:

This report was developed by the Radiation Advisory Committee (RAC) of the Science Advisory Board (SAB) in response to a request to SAB Director, Dr. Donald G. Barnes from Ms. E. Ramona Trovato, Director of EPA's Office of Radiation and Indoor Air (ORIA). Ms. Trovato requested that the Committee review technical aspects of the Multi-Agency Radiation Survey and Site Investigation Manual (MARSSIM) draft dated December 1996 and respond to the following questions in its review:

a) Is the overall approach to the planning, data acquisition, data assessment, and data interpretation as described in the MARSSIM technically acceptable?

b) Are the methods and assumptions for demonstrating compliance with a dose- or risk-based regulation technically acceptable?

c) Are the hypotheses and statistical tests and their method of application appropriate?

For this review, the RAC formed the MARSSIM Review Subcommittee (the Subcommittee). Because of the inter-agency nature of MARSSIM, invitations were extended to advisory committees of other Federal agencies to provide liaisons to the Subcommittee, as a result of which a member of the U.S. Department of Energy's Environmental Management Advisory Board agreed to participate in this review. The Subcommittee met on January 22-23, 1997 and June 17-18, 1997, at which times it was briefed by and had discussions with members of the multi-agency technical working group. In addition, the Subcommittee conducted a public teleconference on July 21, 1997.

MARSSIM was developed collaboratively by four Federal agencies having authority for control of radioactive materials: U.S. Department of Defense (DOD), U.S. Department of Energy (DOE), U.S. Environmental Protection Agency (EPA), and U.S. Nuclear Regulatory Commission (NRC). The Subcommittee was very impressed with the collaboration demonstrated by these agencies, commissions and departments and commends the agencies and their staff for this effort. The product of this effort (MARSSIM) addresses the need for a nationally consistent approach to conducting radiation surveys of potentially radioactively contaminated sites that are being considered for release to the public. **We strongly encourage the timely completion and publication of MARSSIM, in concert with the establishment of a formal procedure for its future revisions, and it is not our intention that our comments delay its completion.** Most of our comments can be incorporated in the previous document, while some others may require consideration for future documents. We strongly encourage these Federal agencies to continue to work together to address issues related to radiation survey and site investigation.

The attached report addresses the charge and elaborates upon significant issues related to technical aspects of MARSSIM. In particular, this report focuses on specific technical issues that the Subcommittee felt would be most likely to require the attention of the Multi-Agency workgroup in order to provide the most comprehensive and technically-supportable basis for the manual. Specifically, the Subcommittee wishes to bring the following major findings and recommendations to your attention:

a) In general, the Subcommittee found that MARSSIM is nearly a finished product. The multi-agency team is commended for its work in addressing the many complex issues involved, resulting in the compilation of an exceptionally well-prepared reference which is technically sound and which will be a useful tool for guiding final status surveys. The document provides generally consistent and explicit guidance for planning and conducting radiation surveys for the decommissioning of radiologically contaminated sites.

b) MARSSIM should discuss its rationale for limiting its scope to guidance for contaminated surface soils and building surfaces. Furthermore, it should more clearly state that radioactive contamination of subsurface soil, surface water, and ground water are explicitly excluded from its coverage. The document should include some discussion of why these particular media were not included, the potential for incorrect decisions if they are not evaluated and the plans, if any, to cover them in the future. Also, MARSSIM should discuss the extent to which it is necessary to evaluate scenarios under which subsurface contamination might be expected to contribute to surface contamination in the future, and how this affects the decision of whether the site meets release criteria.

c) Descriptions of field measurement methods, instruments, and operating procedures in MARSSIM are technically sound but incomplete. Some additions, clarifications, and corrections are noted in our report. MARSSIM should provide guidance for the development of standardized procedures, including a list of considerations for designing site-specific surface-soil sampling and preparation methods so as to ensure that samples will be representative of the materials of concern in deriving the derived concentration guideline levels (DCGLs) for the site.

d) Descriptions of the selection and operation of radiation detection instruments for laboratory analyses are technically sound and represent standard practice but may not be state-of-the-art. MARSSIM should standardize the level of detail used in its presentation of this material and should also provide information on the planned scope and current status of plans to prepare a manual on Multi-Agency Radiological Laboratory Analytical Protocols (MARLAP). MARLAP may be a more appropriate forum than MARSSIM in which to provide more thorough in-depth guidance to the user on the selection and operation of laboratory instrumentation.

e) The Subcommittee believes that it is critically important that the assumptions and procedures used in MARSSIM to make comparisons with the DCGLs match those used in defining the DCGLs. For example, if a DCGL for soil is derived from a dose limit or risk criterion by assuming that a receptor ranges over a certain area on a random basis, then the same area should be used for spatial averaging in the MARSSIM statistical analyses. Such averaging is usually performed from the standpoint of potential human receptors. The manual should note that different spatial and temporal scales of averaging will be necessary if dose- and risk-based criteria are applied to components of the ecosystem other than humans for derivation of a DCGL. This recommendation assumes that the DCGL is derived in a manner appropriate for characterizing human and/or ecological exposures likely to occur at the site under investigation.

f) Although MARSSIM is applicable to the majority of contaminated sites, there appear to be cases that MARSSIM, as currently written, would have trouble addressing. These include: 1) cases dealing with the release of sites that had been contaminated with naturally occurring radionuclides and in which the DCGL is contained within the ambient (background) analyte variability, and 2) cases in which a reference background cannot be established. The Subcommittee recommends that future revisions of MARSSIM provide guidance to the user regarding appropriate choices when such conditions are encountered. For example, the null hypothesis might be redefined to be that the distribution of site radioactivity is no different from that at the reference site or than ambient radioactivity in general.

g) MARSSIM properly warns the user that the DCGL is not free of error and that the uncertainty associated with this quantity may be considerable if derived using generic assumptions and parameter values. However, its discussion of this issue is relegated to an appendix. This important aspect, together with an expanded discussion of its implications for the release decision, needs to be disclosed more prominently in the text of the main document. It is clearly undesirable to design a survey around a DCGL that may not be relevant to the actual conditions at a site, such that actual exposures, doses, and risks would be largely different than those used to derive the generic DCGL. Consequently, MARSSIM should more strongly encourage the user to examine critically the assumptions made in any model used to derive DCGLs for a site in order to determine whether application of site-specific information and parameters would result in significant modifications to the proposed DCGL, or whether development of a site-specific model would be warranted in order to obtain a DCGL that is more relevant to the human and ecological exposure conditions prevailing at the site.

h) In MARSSIM, the preferred null hypothesis is that a survey unit is not ready for release and the information gathered must be sufficient, with a high degree of confidence, to accept the alternative hypothesis (i.e., that the unit meets the release criteria). Furthermore, MARSSIM discusses in detail two non-parametric procedures, the Wilcoxon Rank-Sum test and the Sign test, for testing this hypothesis. However, MARSSIM allows more flexibility in defining the null hypothesis and in choosing statistical analysis methods to test that hypothesis than may be readily apparent to most readers. The existence of this flexibility needs to be more clearly stated and the criteria for selecting among potentially applicable tests need to be described.

i) MARSSIM’s discussion about the mean and median should be revised in order to ensure that the correct statistical parameter is used to compare concentrations in the survey area to those in the reference area. The target statistic for any exposure assessment should be the arithmetic mean concentration for a defined area, together with the uncertainty associated with the estimate of the mean. For a normally distributed population, the mean and the median are identical in value. However, when the distribution of sample evidence is moderately to highly skewed, then non-parametric statistical techniques cannot be used to determine the uncertainty associated with the estimate of the arithmetic mean, and the median of such a sample set will underestimate the true arithmetic mean of surface contamination. The majority of soil sampling programs usually reveal highly skewed distributions. Therefore, the Wilcoxon Rank-Sum test and the Sign test, which are appropriate for testing differences in median concentrations, may not be appropriate to test for differences in mean concentrations.

j) The guidance provided by MARSSIM may introduce an additional measure of conservatism in the process of setting and determining compliance with radiation cleanup standards, compounding the conservatism already likely to occur in developing default DCGLs. Release decisions may be biased correspondingly. MARSSIM should include a qualitative summary of any biases that may result from its assumptions and policy choices, and recommend that the planning team be similarly revealing when developing a site-specific survey design.

Finally, we offer the following comment on an issue that was outside the scope of our charge but that we felt was important to bring to your attention:

k) DCGLs are critical for determining the acceptability of residual levels of radioactivity remaining after a site has been remediated. The Subcommittee suggests that the various <u>approaches</u> proposed for derivation of DCGLs (not the individual site-specific DCGLs) be reviewed and evaluated. This evaluation can be performed by an interagency group and by the EPA/SAB. This evaluation should focus on the strengths and weaknesses of current methodologies and opportunities to refine generic DCGLs with improved site-specific models and data. This review is important but outside the current scope of the SAB/RAC review of MARSSIM *per se.*

We would like to again commend the multi-agency approach used so successfully to produce MARSSIM and to encourage the timely completion and publication of the document. We believe that all of the above recommendations except for items f) and k) can be incorporated into the document at this time, and the remaining two items can be addressed in the future. We strongly encourage the continuation of this multi-agency approach to plan for future revisions to MARSSIM as well as for the development of additional radiation survey manuals, such as for subsurface soils, ground water, and sewers. It would also be beneficial to apply this successful interagency approach to the preparation of other manuals such as on site stabilization, decommissioning techniques, and standardized sampling procedures for various media.

The RAC and its Subcommittee appreciate the opportunity to provide this report to you and we hope that it will be helpful. We look forward to your response to this report in general, and to the comments and recommendations in this letter in particular.

Sincerely,

/signed/
Dr.Genevieve M. Matanoski, Chair
Science Advisory Board

/signed
Dr. James E. Watson, Jr., Chair
Radiatio Advisory n Committee and
MARSSIM Review Subcommittee
Science Advisory Board

NOTICE

This report has been written as a part of the activities of the Science Advisory Board, a public advisory group providing extramural scientific information and advice to the Administrator and other officials of the Environmental Protection Agency. The Board is structured to provide a balanced, expert assessment of scientific matters related to problems facing the Agency. This report has not bbeen reviewed for approval by the Agency; hence, the comments of this report do not necessarily reflect the views and policies of the Environmental Protection Agency or of other Federal Agencies. Any mention of trade names or commercial products does not constitute endorsement or recommendation for use.

ABSTRACT

The EPA Science Advisory Board's (SAB) Radiation Advisory Committee (RAC) / MARSSIM Review Subcommittee (the Subcommittee) reviewed technical aspects of the Multi-Agency Radiation Survey and Site Investigation Manual (MARSSIM) (12/96). The reviewed document was developed collaboratively by four Federal agencies, departments and commissions having authority for control of radioactive materials: Department of Defense, Department of Energy, Environmental Protection Agency, and Nuclear Regulatory Commission. MARSSIM addresses the need for a nationally consistent approach to conducting radiation surveys of potentially radioactively contaminated sites that are being considered for release to the public. A condition of release is a demonstration that residual radioactivity levels do not exceed a specified risk or dose level, also known as a release criterion. MARSSIM provides guidance to users assessing the survey results for surface soils and building surfaces. The Subcommittee concluded that MARSSIM needed to more clearly emphasize that its scope is limited to guidance for surficial media and not to radioactive contamination of any other media. The Subcommittee found that descriptions of field and laboratory measurement methods, instruments and operating procedures in MARSSIM were generally technically sound although somewhat incomplete or out-of-date. The Subcommittee stressed that MARSSIM needed to revise its guidance on the use of the median in place of the mean to represent the average contaminant level in an area and to more clearly state the user's flexibility in selecting statistical methods for evaluating analytical data against the release criterion. The Subcommittee recommended that MARSSIM provide a prominent discussion of the uncertainties and level of conservatism inherent in the setting of the release criterion.

Key Words: Cleanup Standards, Environmental Radiation, Nuclear Facilities, Environmental Quality, Radionuclide Cleanup, Radiological Characterization of Site

SCIENCE ADVISORY BOARD
RADIATION ADVISORY COMMITTEE (RAC)
MULTI-AGENCY RADIATION SURVEY AND SITE INVESTIGATION MANUAL REVIEW SUBCOMMITTEE (MARSSIMRS)

Chair

Dr. James E. Watson, Jr., Professor, Department of Environmental Sciences and Engineering, University of North Carolina at Chapel Hill, NC

Members

Dr. William Bair, (Retired) Former Manager, Life Sciences Center, Battelle Pacific Northwest Laboratory, Richland, WA

Dr. Stephen L. Brown, Director, R2C2 (Risks of Radiation and Chemical Compounds), Oakland, CA (Statistics Workgroup Coordinator)

Dr. June Fabryka-Martin, Staff Scientist, Chemical Science and Technology Division, Los Alamos National Laboratory, Los Alamos, NM

Dr. Thomas F. Gesell, Professor of Health Physics and Director, Technical Safety Office, Idaho State University, Pocatello, ID

Dr. F. Owen Hoffman, President, SENES Oak Ridge, Inc., Center for Risk Analysis, Oak Ridge, TN

Dr. Janet Johnson, Senior Radiation Scientist, Shepherd Miller, Inc., Fort Collins, CO (Monitoring Workgroup Coordinator)

Dr. Bernd Kahn, Professor, School of Nuclear Engineering and Health Physics, and Director, Environmental Resources Center, Georgia Institute of Technology, Atlanta, GA

Dr. Ellen Mangione, M.D., M.P.H., Director, Disease Control and Environmental Epidemiology Division, Colorado Department of Health, Denver, CO

Dr. Paul J. Merges, Chief, Bureau of Pesticides and Radiation, Division of Solid and Hazardous Materials, New York State Department of Environmental Conservation, Albany, NY (Integration Workgroup Coordinator)

Consultants

Dr. Michael E. Ginevan, M.E. Ginevan & Associates, Silver Spring, MD

Dr. David G. Hoel, Chairman and Professor, Department of Biometry & Epidemiology, Medical University of South Carolina, Charleston, SC

Dr. David E. McCurdy, Chief Scientist, Yankee Atomic Electric Company, Bolton, MA

Dr. Frank L. Parker, Vanderbilt University, Nashville, TN [Liaison from Environmental Management Advisory Board (EMAB), U.S. Department of Energy (DOE)]

Science Advisory Board Staff

Dr. K. Jack Kooyoomjian, Designated Federal Official, U.S. EPA, Science Advisory Board (1400), 401 M Street, SW, Washington, DC 20460

Mrs. Diana L. Pozun, Staff Secretary, U.S. EPA, Science Advisory Board (1400), 401 M Street, SW, Washington, DC 20460

TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

List of Appendices:

1. EXECUTIVE SUMMARY

The EPA Science Advisory Board's (SAB) Radiation Advisory Committee (RAC) / MARSSIM Review Subcommittee (the Subcommittee) reviewed technical aspects of the Multi-Agency Radiation Survey and Site Investigation Manual (MARSSIM) (12/96). The reviewed document was developed collaboratively by four Federal agencies, departments and commissions having authority for control of radioactive materials: Department of Defense, Department of Energy, Environmental Protection Agency, and Nuclear Regulatory Commission. MARSSIM addresses the need for a nationally consistent approach to conducting radiation surveys of potentially radioactively contaminated sites that are being considered for release to the public. A condition of release is a demonstration that residual radioactivity levels do not exceed a specified risk or dose level, also known as a release criterion. MARSSIM provides guidance to users performing and assessing the results of such a demonstration for surface soils and building surfaces.

The Subcommittee was requested by the Agency's Office of Radiation and Indoor Air (ORIA) to respond to the following charge in its review:

a) Is the overall approach to the planning, data acquisition, data assessment, and data interpretation as described in the MARSSIM technically acceptable?

b) Are the methods and assumptions for demonstrating compliance with a dose- or risk-based regulation technically acceptable?

c) Are the hypotheses and statistical tests and their method of application appropriate?

The Subcommittee's full report addresses this charge and elaborates upon significant issues related to technical aspects of MARSSIM. The report focuses on specific technical issues that the Subcommittee felt would be most likely to require the attention of the multi-agency workgroup in order to provide the most comprehensive and technically-supportable basis for the manual. The Subcommittee made the following major findings and recommendations:

a) In general, the Subcommittee found that MARSSIM is nearly a finished product. The multi-agency team is commended for its work in addressing the many complex issues involved, resulting in the compilation of an exceptionally well-prepared reference which is technically sound and which will be a useful tool for guiding final status surveys. The document provides generally consistent and explicit guidance for planning and conducting radiation surveys for the decommissioning of radiologically contaminated sites.

b) MARSSIM should discuss its rationale for limiting its scope to guidance for contaminated surface soils and building surfaces. Furthermore, it should more clearly state that radioactive contamination of subsurface soil, surface water, and ground water are explicitly excluded from its coverage. The document should include some discussion of why these particular media were not included, the potential for incorrect decisions if they are not evaluated and the plans, if any, to cover them in the future. Also, MARSSIM should discuss the extent to which it is necessary to evaluate scenarios under which subsurface contamination might be expected to contribute to surface contamination in the future, and how this affects the decision of whether the site meets release criteria.

c) Descriptions of field measurement methods, instruments, and operating procedures in MARSSIM are technically sound but incomplete. Some additions, clarifications, and corrections are noted in our report. MARSSIM should provide guidance for the development of standardized procedures, including a list of considerations for designing site-specific surface-soil sampling and preparation methods so as to ensure that samples will be representative of the materials of concern in deriving the derived concentration guideline levels (DCGLs) for the site.

d) Descriptions of the selection and operation of radiation detection instruments for laboratory analyses are technically sound and represent standard practice but may not be state-of-the-art. MARSSIM should standardize the level of detail used in its presentation of this material and should also provide information on the planned scope and current status of plans to prepare a manual on Multi-Agency Radiological Laboratory Analytical Protocols (MARLAP). MARLAP may be a more appropriate forum than MARSSIM in which to provide more thorough in-depth guidance to the user on the selection and operation of laboratory instrumentation.

e) The Subcommittee believes that it is critically important that the assumptions and procedures used in MARSSIM to make comparisons with the DCGLs match those used in defining the DCGLs. For example, if a DCGL for soil is derived from a dose limit or risk criterion by assuming that a receptor ranges over a certain area on a random basis, then the same area should be used for spatial averaging in the MARSSIM statistical analyses. Such averaging is usually performed from the standpoint of potential human receptors. The manual should note that different spatial and temporal scales of averaging will be necessary if dose- and risk-based criteria are applied to components of the ecosystem other than humans for derivation of a DCGL. This recommendation assumes that the DCGL is derived in a manner appropriate for characterizing human and/or ecological exposures likely to occur at the site under investigation.

f) Although MARSSIM is applicable to the majority of contaminated sites, there appear to be cases that MARSSIM, as currently written, would have trouble addressing. These include: 1) cases dealing with the release of sites that had been contaminated with naturally occurring radionuclides and in which the DCGL is contained within the ambient (background) analyte variability, and 2) cases in which a reference background cannot be established. The Subcommittee recommends that future revisions of MARSSIM provide guidance to the user regarding appropriate choices when such conditions are encountered. For example, the null hypothesis might be redefined to be that the distribution of site radioactivity is no different from that at the reference site or than ambient radioactivity in general.

g) MARSSIM properly warns the user that the DCGL is not free of error and that the uncertainty associated with this quantity may be considerable if derived using generic assumptions and parameter values. However, its discussion of this issue is relegated to an appendix. This important aspect, with an expanded discussion of its implications for the release decision, needs to be disclosed more prominently in the text of the main document. It is clearly undesirable to design a survey around a DCGL that may not be relevant to the actual conditions at a site, such that actual exposures, doses, and risks would be largely different than those used to derive the generic DCGL. Consequently, MARSSIM should more strongly encourage the user to examine critically the assumptions made in any model used to derive DCGLs for a site in order to determine whether application of site-specific information and parameters would result in significant modifications to the proposed DCGL, or whether development of a site-specific model would be warranted in order to obtain a DCGL that is more relevant to the human and ecological exposure conditions prevailing at the site.

h) In MARSSIM, the preferred null hypothesis is that a survey unit is not ready for release and the information gathered must be sufficient, with a high degree of confidence, to accept the alternative hypothesis (i.e., that the unit meets the release criteria). Furthermore, MARSSIM discusses in detail two non-parametric procedures, the Wilcoxon Rank-Sum test and the Sign test, for testing this hypothesis. However, MARSSIM allows more flexibility in defining the null hypothesis and in choosing statistical analysis methods to test that hypothesis than may be readily apparent to most readers. The existence of this flexibility needs to be more clearly stated and the criteria for selecting among potentially applicable tests need to be described.

i) MARSSIM's discussion about the mean and median should be revised in order to ensure that the correct statistical parameter is used to compare concentrations in the survey area to those in the reference area. The target statistic for any exposure assessment should be the arithmetic mean concentration for a defined area, together with the uncertainty associated with the estimate of the mean. For a normally distributed population, the mean and the median are identical in value. However, when the distribution of sample evidence is moderately to highly skewed, then non-parametric statistical techniques cannot be used to determine the uncertainty associated with the estimate of the arithmetic mean, and the median of such a sample set will underestimate the true arithmetic mean of surface contamination. The majority of soil sampling programs usually reveal highly skewed distributions. Therefore, the Wilcoxon Rank-Sum test and the Sign test, which is appropriate for testing differences in median concentrations, may not be appropriate to test for differences in mean concentrations.

j) The guidance provided by MARSSIM may introduce an additional measure of conservatism in the process of setting and determining compliance with radiation cleanup standards, compounding the conservatism already likely to occur in developing default DCGLs. Release decisions may be biased correspondingly. MARSSIM should include a qualitative summary of any biases that may result from its assumptions and policy choices, and recommend that the planning team be similarly revealing when developing a site-specific survey design.

Finally, the Subcommittee offered the following comments on issues that were outside the scope of the charge but that were felt to be important:

k) DCGLs are critical for determining the acceptability of residual levels of radioactivity remaining after a site has been remediated. The Subcommittee suggested that the various <u>approaches</u> proposed for derivation of DCGLs (not the individual site-specific DCGLs) be reviewed and evaluated. This evaluation can be performed by an interagency group and by the EPA/SAB. This evaluation should focus on the strengths and weaknesses of current methodologies and opportunities to refine generic DCGLs with improved site-specific models and data. This review is important but outside the current scope of the SAB/RAC review of MARSSIM *per se.*

l) The Subcommittee strongly encouraged the continuation of this multi-agency approach to plan for future revisions to MARSSIM as well as for the development of additional radiation survey manuals, such as for subsurface soils, ground water, and sewers. It would also be beneficial to apply this successful interagency approach to the preparation of other manuals such as on site stabilization, decommissioning techniques, and standardized sampling procedures for various media.

2. INTRODUCTION

2.1 Overview of the Multi-Agency Radiation Survey and Site Investigation Manual

A Multi-Agency Radiation Survey and Site Investigation Manual (MARSSIM) (U.S. EPA et al., 1996) is being developed collaboratively by four Federal agencies, departments and commissions having authority for control of radioactive materials: U.S. Department of Defense (DOD), U.S. Department of Energy (DOE), U.S. Environmental Protection Agency (EPA), and U.S. Nuclear Regulatory Commission (NRC). The December 1996 draft MARSSIM report reviewed by the EPA's Science Advisory Board (SAB) was prepared by a multi-agency technical working group composed of representatives from these four Federal agencies, departments, and commissions. When finalized, MARSSIM will be a multi-agency consensus document.

MARSSIM addresses the need for a nationally consistent approach to conducting radiation surveys of potentially radioactively contaminated sites that are being considered for release to the public. A condition of release is a demonstration to the responsible Federal or state agency that any residual radioactivity levels do not exceed a specified risk or dose level, also known as a release criterion, established by the responsible agency. MARSSIM assists site personnel and others in performing and assessing the results of such a demonstration for surface soils and building surfaces. The guidance provided in MARSSIM is intended to be not only scientifically rigorous but also sufficiently flexible to be applied to a diversity of sites at different stages of the cleanup process.

Guidance is provided in MARSSIM on historical site assessment, preliminary survey considerations, survey planning and design, field measurement methods and instrumentation, sampling and preparation for laboratory measurements, interpretation of survey results, and quality assurance and quality control measures. Survey types considered include scoping, characterization, remedial action support, and final status surveys. For each type of survey, guidance is provided on survey design, conducting the survey, evaluating results, and documentation. Results of the final status survey are used to determine whether or not the release criterion has been met. Statistical tests are presented for use in the decision-making process.

MARSSIM notes several areas that are beyond its scope. These areas include translation of dose or risk standards into radionuclide-specific concentrations; demonstration of compliance with ground water or surface water regulations; management of vicinity properties not under government or licensee control; surveys of other contaminated media (such as subsurface soil, building materials, and ground water); and the release of contaminated components and equipment.

2.2 Charge to the SAB

The multi-agency technical working group that prepared the draft MARSSIM agreed to request review of this draft by the SAB through its Radiation Advisory Committee (RAC). This request was submitted to Dr. Donald G. Barnes, the SAB Director, by Ms. E. Ramona Trovato, Director of EPA's Office of Radiation and Indoor Air (ORIA) in a memo dated July 30, 1996. Ms. Trovato requested that the Committee respond to the following questions in its review:

a) Is the overall approach to the planning, data acquisition, data assessment, and data interpretation as described in the MARSSIM technically acceptable?

b) Are the methods and assumptions for demonstrating compliance with a dose- or risk-based regulation technically acceptable?

c) Are the hypotheses and statistical tests and their method of application appropriate?

Because of the inter-agency nature of MARSSIM, Ms. Trovato also noted that it would be appropriate to involve persons from advisory committees of the other Federal agencies.

2.3 SAB Review Procedure

The primary review document is the December 1996 draft report, titled "Multi-Agency Radiation Survey and Site Investigation Manual (MARSSIM)" (U.S. EPA et al., 1996). Additional background documentation was provided by ORIA in support of this review.

For this review, the RAC formed the MARSSIM Review Subcommittee (the Subcommittee). Invitations were extended to advisory committees of other Federal agencies, departments, and commissions to provide liaisons to the Subcommittee, as a result of which a member of DOE's Environmental Management Advisory Board agreed to participate in this review. Three workgroups were formed within the Subcommittee to focus on areas of the report dealing with a) the integration of the MARSSIM process, b) radiation measurement activities, and c) statistical analyses of the measurement data. The Subcommittee met on January 22-23, 1997 and June 17-18, 1997, at which times it was briefed by and had discussions with members of the multi-agency technical working group. In addition, the Subcommittee conducted a public teleconference meeting on July 21, 1997 to reach closure on this draft report.

3. ASSESSMENT OF MARSSIM's OVERALL APPROACH

3.1 Response to Charge a)

The charge from ORIA's director regarding MARSSIM's overall approach was:

Charge a) Is the overall approach to the planning, data acquisition, data assessment, and data interpretation as described in the MARSSIM technically acceptable?

MARSSIM was developed collaboratively by four Federal agencies, departments, and commissions having authority for control of radioactive materials: Department of Defense (DOD), Department of Energy (DOE), Environmental Protection Agency (EPA), and Nuclear Regulatory Commission (NRC). The Subcommittee was impressed with the collaboration demonstrated by these agencies, departments, and commissions, and commends these organizations and their staff for this effort. The product of this effort (MARSSIM) addresses the need for a nationally consistent approach to conducting radiation surveys of potentially radioactively contaminated sites that are being considered for release to the public. With the incorporation of the comments in our report, the Subcommittee feels that the overall approach to the planning, data acquisition, data assessment, and data interpretation as described in MARSSIM will be technically acceptable and that MARSSIM will fulfill its stated objectives.

In its deliberations on this charge, the Subcommittee concluded that MARSSIM should more clearly emphasize that its scope is limited to guidance for contaminated surface soils and building surfaces. Other considerations not addressed in MARSSIM include the radioactive contamination of subsurface soil, surface water, and ground water. The document should include some discussion of why these particular media were not included and the plans, if any, to cover them in the future. Also, the Subcommittee judged it important that MARSSIM at least mention scenarios under which subsurface contamination might contribute to surface contamination in the future. Detailed discussion of each of these points is provided below.

We encourage the timely completion and publication of MARSSIM, and it is not our intention that our comments delay its completion. Most of our comments can be incorporated in the previous document, while some others will require consideration for future documents. We encourage these Federal agencies to continue to work together to address issues related to radiation survey and site investigation.

3.2 MARSSIM's Scope of Coverage

3.2.1 Relationship to Federal Regulations

The MARSSIM brings together, for the first time of which the Subcommittee is aware, DOE, EPA, NRC, and DOD with a common method of site surveys and investigations for surficial contamination. Adopting MARSSIM will mean that surveys done for any of the participating organizations will be immediately transparent to all of the others. The information in MARSSIM's Appendix F is particularly useful for understanding the complex interrelationships between radiological surveys undertaken following MARSSIM guidelines and those undertaken as part of CERCLA and RCRA cleanup actions. In fact, if the site has significant chemical contamination that might limit its suitability for unrestricted release, then it may be efficient to coordinate investigations and sampling with the corresponding CERCLA Remedial Investigation/Feasibility Study (RI/FS) or closure surveys.

On the other hand, because of its close tie to various regulatory programs, MARSSIM also has the challenge and responsibility to provide information that is consistent with the current regulatory picture. Because this aspect is somewhat of a moving target, we recommend that references not be made to regulations in draft form, and that such references in MARSSIM instead be revised to refer to the regulatory responsibilities of the agencies. For example, the discussion on page 1-1, lines 8-14, could be revised as follows (proposed changes are underlined):

> "The Environmental Protection Agency (EPA), the Nuclear Regulatory Commission (NRC), and the Department of Energy (DOE) <u>are charged with the preparation of</u> regulations for the release of certain categories of radioactively contaminated sites following such cleanup.... Some states <u>may prepare</u> similar rules that will apply to sites under their control."

It needs to be emphasized early in MARSSIM that several critical aspects of the cleanup process are not within its scope. For example, MARSSIM provides guidance on verifying cleanup but is not intended to provide guidance on methods for cleanup activities nor on methods for setting cleanup standards. Furthermore, MARSSIM should note for its users that compliance with the Derived Concentration Guideline Level (DCGL) is only one of the considerations for release of a site for unrestricted use. Other considerations not addressed in MARSSIM include the radioactive contamination of subsurface soil, surface water, or ground water. Furthermore, if the release criterion is supposed to be ultimately risk-based, how will the risks of residual chemicals be factored into the decision about release criteria for radionuclides (see MARSSIM page 3-5)? If a site has mixed hazardous and radiological contamination, their combined risk will need to be considered.

MARSSIM should discuss the extent to which it is necessary to evaluate scenarios under which subsurface contamination might be expected to contribute to surface contamination in the future, and how this affects the decision of whether the site meets release criteria. For example, cleanup under 40 CFR 192 (Uranium Mill Tailing Radiation Control Act, UMTRCA) standards allows a higher concentration of radionuclides to be left below the surface, which may well be exposed following a large flooding or erosion event.

3.2.2 MARSSIM as a Guidance Document

A major concern of the Subcommittee is the use to which MARSSIM will be put, i.e., is there a risk that this document will be used by regulatory agencies as *de facto* regulation rather than guidance, discouraging flexibility in approaches and precluding use of creative and reasonable site-specific measurement methodology? This concern is particularly applicable to sites where the radiological constituents may be present at levels not distinguishable from background. Consequently, the Subcommittee suggests that Section 2.6 in MARSSIM either begins with a paragraph, or else adds a new section (e.g., a new Section 2.6.5), that emphasizes the function of the document for guidance and that offers the opportunity for the surveyor and investigator to provide an alternative plan to the regulatory agency. This plan has to meet the agency's criteria for acceptable and pertinent information, but can be adapted to the specific circumstances of the site.

While in the realm of policy options, the Subcommittee noted a related concern with the question of sufficiency, i.e., will use of methods described in MARSSIM be considered sufficient to show compliance with regulatory requirements? As an extension of this issue, if a scientific peer review has concluded that the methods in MARSSIM are sufficient on technical grounds to decide whether or not a particular site meets its DCGL criterion, then to what extent will a regulatory agency be bound to accept findings based on surveys performed strictly according to MARSSIM guidance?

The document is confusing by the extent to which various terms and requirements have legal or regulatory connotations, as opposed to common convention. For example, MARSSIM should explain where the concepts of DCGL—and its subsets $DCGL_W$ and $DCGL_{EMC}$—come from. References for the classification of areas as Class 1, Class 2 and Class 3, the use of a graded approach, and the use of the surrogate approach or the unity rule should also be provided. MARSSIM should also clearly distinguish between actions that are merely good practice and those that are necessary to ensure defensible data. Examples of the latter case are maintenance of documentation on sample chain of custody and use of calibrated instrumentation.

3.2.3 Use and Citations of References

The reader's confidence that the document presents up-to-date information is eroded by the extent to which supporting references are outdated, inappropriate, or not cited at all. Although MARSSIM may well be the best compilation of cleanup documents available in the U.S., its citations and references on industry standards and instrumentation could be improved. In particular, references to publications by the American National Standards Institute (ANSI) and the American Society of Testing and Materials (ASTM) should be updated because these are revised on a five-year cycle. Also, many of the references cited in the report do not appear in the list of references, and vice versa as well. Numerous examples of these shortcomings are listed in our detailed comments submitted separately to ORIA (Kooyoomjian, 1997).

There is also a heavy reliance on the unpublished literature to support the information in MARSSIM, including contractor reports, reports at meetings, and even draft documents. In many other instances, information is provided without identifying any reference. In some cases there may be no choice; and we recognize that government or contractor reports, while not always easily accessible, may nonetheless be the best or most appropriate information sources. In other cases, however, equivalent or better sources can be found in the conventionally published literature. For example, there are several sections on instrumentation in the MARSSIM draft but no citation of any of the standard nuclear instrumentation texts except for a citation to Knoll (1979) regarding propagation of error, for which a statistics book might have been a better source. Even then, Knoll is not listed in the reference section nor is 1979 the most recent edition (it is 1989). The credibility of the MARSSIM report would be greatly enhanced by a strong reference section that includes recognized, readily accessible literature wherever possible.

3.2.4 Application to Surface Soil and Building Surfaces

In the view of the Subcommittee, MARSSIM is largely intended to guide the user in the acquisition and assessment of radiological data relevant to the decision about releasing the site. Of course, if the site fails, remediation is probably indicated, and the survey can help identify the areas needing such work. But the survey should not be designed to tell a site manager whether or not 10 pCi/g of radium in sewer sludge is acceptable because such a decision would require developing a specific scenario for exposure to the sludge so that dose or risk can be calculated. The same may be true for more dispersed radioactivity, such as a layer of higher activity material in soil at some depth beneath the surface. Although it may be important to warn MARSSIM users about such possibilities, MARSSIM should not attempt to define methods for every such contingency.

In support of the Subcommittee's impression of a limited scope of coverage, Chapter 1 of MARSSIM clearly establishes its scope as providing guidance for contaminated surface soils and building surfaces. In this regard, the title of the document is misleading; it would be more appropriately called MARSSSIM—Multi-Agency Radiological Surficial Survey and Site Investigation Manual— because it only addresses "surficial" characterization of the top 15 cm of soils (see MARSSIM p. 2-31, lines 663-664 defining "site boundary") and of building surfaces. The contents of MARSSIM should be consistent with its stated scope of coverage throughout, such that sections not clearly falling within the scope should be omitted or moved to an appendix. Examples include guidance on sampling surface water, ground water, sediment, vegetation, aerosols, and subsurface soils. If these sections are left in the report in any form, then they raise the question as to whether or not the DCGLs established for surface soils and building surfaces also apply to these media. It would be useful and appropriate to add an explanation in Chapter 1 as to why MARSSIM is limited to surface soils and buildings, and what the user should do for other types of contaminated media.

MARSSIM should clarify what constitutes "surface soil" by citing a regulatory-based definition (e.g., the appropriate section in 40 CFR Part 192). In one section of MARSSIM as well as the glossary, surface soil is described as being the top 15 cm without indicating whether or not this particular depth is an arbitrary decision or based on regulatory requirements. In addition, MARSSIM never makes clear the extent to which the user should be attempting to characterize contamination over this entire depth. Numerous consequences would result if this is so. For example, the quantity of sample collected in some cases would be orders of magnitude larger than needed for an analytical measurement. Should it matter that contamination limited to the top 1-cm would possibly be diluted to below the DCGL if it were mixed to a depth of 15 cm during the sampling process? Should pavement samples also be collected to a depth of 15 cm? How should cobbly or skeletal soils be sampled? The importance of homogenizing samples before removing an aliquot for analysis should be discussed, particularly if replicate analyses are part of the standard quality control protocol for judging data reproducibility (see Sections 3.5.4 and 3.5.5 of this report). It is also critical to examine how the definition of surface soil relates to how the DCGL is established from pathway modeling (see Section 4.2 of this report). MARSSIM should provide guidance to the user on the development of standardized methods for field measurements and for sample collection and processing in order to address these areas of concern (see Section 3.3).

3.2.5 Application to Subsurface Soil and Other Environmental Media

In addition to more explicit discussion of topics that are included in the present scope of MARSSIM (see Section 3.2.4 above), the manual should also be more explicit in an early section in its identification of topics that are outside its scope. Furthermore, it should ensure that the manual's contents stay consistent with these limits. In this proposed section, the document should include discussion of the rationale for excluding coverage of these specific topics and the plans, if any, to cover them in the future, or else where to find guidance on them now. Examples of environmental media outside the scope of MARSSIM include subsurface soils (presumably meaning soils more than 15 cm beneath the surface), ground water, and buildings (beyond surface contamination). As presented by the MARSSIM co-authors in their discussions with the Subcommittee, the rationale for excluding these media includes such sound reasons as:

a) contamination is limited to surface soil in the majority of the sites (80-90%) to which the MARSSIM approach would be applied;

b) as a consequence of this fact, existing computer models for dose assessment generally only consider surface soils;

c) MARSSIM was written in support of cleanup rulemaking for which most of the supporting analyses have been limited to contaminated surface soils and building surfaces; and

d) a limited scope was necessary in order to ensure the development and issuance of a useful product—with multi-agency consensus—within a reasonable time.

We recommend that MARSSIM contain wording similar to that above in its introductory material.

MARSSIM should also warn the user that its application may be sufficient only in those cases where the surface soil contains the majority of the overall inventory of contamination. If contamination of other environmental media is a possibility, then following MARSSIM's guidance for surveying a site may be inadequate as a basis for determining the site's suitability for release. Consequently, the document should more clearly state that its domain for application is for cases where the surface soil is the dominant source of human and ecological exposure, dose, and risk. But even in this case, it may be appropriate to advise the user that some subsurface sampling may be required in order to prove that the contamination was limited to the surface and that any subsurface radiological contamination will not be deposited on the surface at some future date (e.g., through being exposed via erosion or excavation).

3.3 Planning Guidance

3.3.1 Overall Approach to Planning the Surveys

The Road Map provided in MARSSIM following the appendices is very useful for understanding the overall MARSSIM approach to site characterization. It should be moved to Chapter 2, and the reader should be strongly urged to review it carefully as an introduction and integration of the MARSSIM process. Provision of an abstract and Executive Summary would also help the reader understand the overall approach.

MARSSIM guidance is unclear as to when the results from the scoping and characterization scans or surveys may also be applicable to the final status survey. MARSSIM specifically states that in order to use scoping and characterization data, they must be of adequate quality to meet the DQOs. (For example, MARSSIM references as to the applicability of characterization and scoping survey data to the final status survey occur on page 2-25, lines 487-490; page 5-3, lines 88-91 and 99-101; and page 5-10, lines 287-289). Further clarification of the issue of data applicability, including some examples, might be useful to the reader. Also, as noted during presentations to the Subcommittee at its January 22, 1997 meeting, MARSSIM is intended to apply to the final site status survey and not to scoping or characterization surveys undertaken for the specific purpose of planning remedial action. The latter objective could involve different guidelines for sampling strategies and analyses. This fact should be strongly emphasized in the discussion of MARSSIM's scope.

3.3.2 Public Involvement

The MARSSIM recommends including stakeholder group representatives on the planning team. Thus, the public will be involved in the early planning stages of any initial site evaluations and surveys. How the planning team would be constituted and how it would arrive at consensus on these and other survey design issues are less clear, especially the role of the public stakeholders. In addition, several key aspects of the MARSSIM process are not straightforward and require thoughtful evaluation by the planning team of the various alternatives and their implications for the release decision. For example, the document discusses the potential difficulties in identifying suitable areas for establishing background radiological conditions against which the site's conditions will be compared. It also recognizes that use of default DCGLs and standardized values for Type I and Type II decision errors (α and β, respectively) may not be appropriate for all sites and consequently allows for comparisons with site-specific DCGLs using decision criteria selected by the planning team.

Although the Subcommittee understands that such issues cannot be fully discussed in a technical document like MARSSIM, they are important for the public acceptability of its results. We therefore recommend that the document acknowledge the issue of public participation and, to the extent possible, reference activities within the organizations that deal with public involvement in the design and implementation of radiation surveys for site release decisions (e.g., in its section 3.2).

4. DATA ACQUISITION AND ASSESSMENT

4.1 Response to Charge b)

The charge from ORIA's director regarding data acquisition and assessment was:

Charge b) Are the methods and assumptions for demonstrating compliance with a dose- or risk-based regulation technically acceptable?

Measurement methods applicable to and appropriate for use in demonstrating compliance with the DCGLs are discussed in Chapters 6 and 7 and Appendix H of MARSSIM. Chapter 6 describes methods and instrumentation for collecting data in the field, i.e., direct measurement of radionuclides in surface soils and on structure surfaces and scanning surveys. Chapter 7 addresses issues related to sampling methods and laboratory instrumentation used for analysis of the samples. Appendix H is a comprehensive compilation of descriptions of instruments available for use in demonstrating compliance with the DCGLs. Chapter 9 also deals with an integral part of data acquisition in that it covers quality assurance and quality control measures.

The Subcommittee finds the description of data acquisition methods to be technically acceptable and concludes that MARSSIM is a very good compilation of methods and instruments and will be a useful tool for regulators and the regulated community. However, treatment of field and laboratory operations in MARSSIM could be improved in the following areas:

a) In MARSSIM Section 6.4.1, Direct Measurement Sensitivity, there appears to be an inconsistency between the presentation (and terminology) of Figure 6.2 and the subsequent (pp 6-20 and 6-21) derivation and definitions of the critical level L_C and the *a priori* detection level L_D. The classical approach developed by L. Currie (Currie, 1968) for these two parameters used figures and derivations based on a net signal response ($\mu_s = 0$) distribution together with the standard deviation (σ_0) of the net signal response distribution. It is recommended that Section 6.4.1 be revised to show consistency between the included figure and the derivation of the subsequent instrument detection parameters.

b) Sample collection protocols and analytical techniques are discussed in MARSSIM Chapters 6 and 7 and in Appendix H. The redundancy in coverage between Chapters 6 and 7 should be minimized by careful editing and cross-referencing, by redefining the scope of each chapter, or by combining the two chapters into a single one, resulting in a more concise, user-friendly, and internally consistent document. Standardized nomenclature should be used throughout MARSSIM in references to field and laboratory equipment in order to avoid confusion or ambiguity.

c) In general, MARSSIM contains technically sound descriptions of field measurement methods, instruments, and operating procedures. Some additions, clarifications, and corrections are noted in our report. For example, some sampling methods described in MARSSIM are incomplete. MARSSIM should provide guidance for the development of standardized sampling procedures for surface soils. Standard procedures should be referenced, such as ASTM C998-83, "Standard Method for Sampling Surface Soil for Radionuclides," and ASTM C999-83, "Standard Method for Soil Sample Preparation for the Determination of Radionuclides." However, the user should also be encouraged to design site-specific surface-soil sampling procedures in order to ensure representative samples of the surface materials that are of concern in deriving the DCGLs for the site. MARSSIM should provide a list of considerations for designing these site-specific surface soil sampling methods. For example, the depth to which a sample is taken may affect the measured concentration if the radionuclide is deposited in the top few centimeters. Under some circumstances, averaging over the top 15 cm is appropriate if the exposure pathway of concern is ingestion of food raised in the area, but may underestimate the potential dose if the exposure pathway of concern is soil ingestion or inhalation of resuspended dust. Other considerations might include the size fraction to be collected and/or analyzed and whether vegetation, large gravels or cobbles, or debris should be removed in the field or prior to laboratory analysis.

d) Descriptions of the selection and operation of radiation detection instruments for laboratory analyses are technically sound and represent standard practice but may not be state-of-the-art. MARSSIM should standardize the level of detail used in its presentation of this material and should also provide information on the planned scope and current status of plans to prepare a manual on Multi-Agency Radiological Laboratory Analytical Protocols (MARLAP), which may be a more appropriate forum in which to provide more thorough in-depth guidance to the user on the selection and operation of laboratory instrumentation.

Specific findings and recommendations of the Subcommittee with regard to these aspects of MARSSIM are given in the following sections.

4.2 Organization of Information on Analytical Instruments in MARSSIM

Sample collection protocols and analytical techniques are discussed in MARSSIM Chapters 6 and 7 and in Appendix H. The distinction between the scopes of Chapters 6 and 7 is not altogether clear, and there is some overlap between the two, particularly with respect to radon analyses (MARSSIM Sections 6.6 and 7.4.5). The excessive redundancy in coverage could be minimized by careful editing and cross-referencing, by redefining the scope of each chapter, or by combining the two chapters into a single one, resulting in a more concise, user-friendly, and internally-consistent document. An example of the redundancy can be found in sections on radon measurements (MARSSIM Sections 6.6.1 and 7.4.5), and other examples are mentioned in detailed comments submitted to ORIA separately from this report (Kooyoomjian, 1997).

Standardized nomenclature should be used throughout MARSSIM in references to field and laboratory equipment in order to avoid confusion or ambiguity. All techniques mentioned in the main body of MARSSIM should have corresponding detailed descriptions in Appendix H, but it is difficult to check the extent to which this recommendation has already been met because of the different terminologies that have been used (MARSSIM Tables 6.1 to 6.3, Chapter 6 and 7 text, Table 7.2, Appendix H).

The level of detail provided for the various techniques is uneven, ranging from no discussion (e.g., low-energy x-rays are mentioned in MARSSIM Chapter 6, line 103, but are not subsequently discussed as have been gamma, alpha, and beta radiation) to overly detailed descriptions (e.g., it is inappropriate in MARSSIM Chapter 6 to discuss the energy response of the PIC (line 149); the need for a site-specific calibration curve for NaI(Tl) detectors (line 152); or n and p type configurations (line 391)).

4.3 Guidance on Data Acquisition

4.3.1 Scanning Surveys

Statements should be made in MARSSIM Section 6.4.2.1 (Scanning for Beta and Gamma Emitters) and in Section 6.4.2.2 (Scanning for Alpha Emitters) as to the applicability of a “scanning survey” as a “final status survey.” Scanning surveys are performed to locate radiation levels or radioactivity above an investigation level (i.e., “hot spots”). The terminology “scanning survey” may be confused with the MARSSIM terminology of “scoping survey”, “characterization survey,” and “final status survey.” Maybe the term “instrument scan” rather than “scanning survey” would be a better choice of words since the former is then similar to the terminology “direct measurement.” Both types of measurements (direct and instrument scan) can be used to characterize the site (MARSSIM Sections 6.2 and 6.2.2).

The statistical treatment of "human factors" in scanning field measurements is inadequate. The section focused on the "Poisson Observer" (p. 6-24, line 221) is technically weak. In particular, the concept is discussed to a far greater extent than appears suitable without adequate documentation and references. The problem should be stated in one or two paragraphs, and a human factors efficiency should be proposed. Information concerning the magnitude of this value and the basis of Equation 6-6 should be referenced. Other approaches for dealing with loss of information during a scanning survey may include analogy to manner in which one deals with the loss of information by an instrument as it moves from one location to another, e.g., by time constant or coincidence counting by paired monitors.

4.3.2 Instrument Calibration

MARSSIM should include a recommendation that calibrations of survey and laboratory instruments be linked to a national standard traceable to the National Institute of Standards and Technology (NIST). Such linkage can be established through a NIST secondary laboratory accreditation program for survey meters (e.g., Eisenhower, 1991), or a program sponsored by the Health Physics Society (HPS, no date), or through one of the national measurement assurance programs (MAP) that is linked to NIST for laboratory analyses (e.g., ANSI, 1996).

4.3.3 Background Measurements and Adjustments to Measured Values

A clear distinction needs to be made among instrumental background, method or process blank background, field blank, and environmental background wherever the term "background" is used. For example, the "s+b" subscript for Equation 6-1 is confusing because this summation refers to total count rate in just the survey area and has nothing to do with the count rate in the reference (background) area, while the subscript "b" in the same equation refers specifically to the count rate in the reference area. (MARSSIM Sections 6.2.5, 6.2.7.1, 9.3.4)

The described approach for the use of surrogates should not be presented as one of deriving a correction factor, but rather an application of a surrogate indicator for the radionuclide of interest. A more realistic example (such as Co-60 and Ni-63 from MARSSIM's Section 4.3.2) should be provided for illustrating this approach (in MARSSIM Section 6.2.7.1) because it does not make sense to expect Co-60 and H-3 to have or to maintain relatively fixed ratios due to their dramatically different chemical behaviors. In addition, guidance should be provided on methods to be used to characterize the fixed ratio and its variability (e.g., number of samples needed).

The derivation of the alpha-scanning equation is provided in Appendix J of MARSSIM. These equations and calculations have been verified by the Subcommittee. However, the basis for the average number of counts expected is confusing. In

Equations 6-7 (page 6-34) and J-3 (page J-2), the average number of counts expected is represented by the term "G E d / 60 v." Although the source activity (G) is defined as the number of decays per minute (dpm) measured over the effective detector area as calculated by the equation on line 48 of Appendix J, the efficiency (E) is not specified as applying to a particular area. The parameter E, which is embedded in the definition of G, should likewise be determined for activity distributed over the detector area.

A related concern is that the derivation of Equation J-5 seems to assume that the dwell time for the entire area of the detector will be the same. This works for rectangular or square detectors but not for circular ones. If the user assumes that the width of the detector in the direction of the scan is equal to the diameter of a circular detector, the true probability of getting a single count at a given level of contamination for a specified scan time will be overestimated.

4.4 Field and Laboratory Instrumentation

4.4.1 MARSSIM's Overview of Instrumentation

Appendix H in MARSSIM provides a useful and informative listing of alternative instrumental techniques for radiation surveys, presented in the format of summary tables and more detailed thumbnail sketches. This appendix is likely to become a frequently consulted reference for decision-makers and planners interested in the advantages and disadvantages of the different techniques. The following suggestions are made to improve the usability of MARSSIM's Appendix H for the intended audience. Subcommittee comments submitted separately to ORIA provide details on specific technical and editorial corrections (Kooyoomjian, 1997).

a) Add a description in the introduction of the appendix that describes each heading used in the detailed descriptions. For example, what factors determine whether or not an instrument is used in the laboratory or field? What does "Secondary Radiation Detected" mean; what is the basis for specifying primary versus secondary? What is the basis of the cost estimates per measurement (e.g., typical quotes from commercial laboratories, vs. inhouse labor rates)? A range of estimates for all costs would be more appropriate than the single value that is given for some of the estimates.

b) The introduction to the appendix would be the place to warn the reader that analytical laboratories need to be contacted for actual cost schedules, and that costs are highly variable depending upon the matrix, turnaround time, sample preparation requirements, the potential for cost savings for large batches, required detection limit, and required level of documentation of quality assurance (QA) or quality control (QC) measures.

c) Consider adding a new heading category, “Advantages and disadvantages,” for the detailed description of each instrumental technique. Such a discussion should be worded in such a way as not to construe endorsement of a particular manufacturer’s product.

d) The detailed descriptions and summary tables could be better organized to provide a more useful thumbnail guide to decision-makers who are planning a survey. The present approach of having multiple sections and multiple summary tables with duplicative information and an uneven level of detail results in much flipping through the pages. A more convenient organization of this information would be to combine these into a single section and a single table, with instruments simply listed alphabetically, and to revise the summary table column headings to focus on those aspects of greatest interest to the survey planner (an example table has been included in the comments transmitted under separate cover to ORIA). The column labeled ‘Remarks’ should be omitted from the summary table because this level of detail is more appropriately discussed in the detailed descriptions.

e) Add descriptions for the additional methods listed on page H-55 (fluorimetry, passivated ion implanted detectors, Cerenkov counter, PERALS scintillation counter, Cd-Zn-Te)

f) There is no specific listing of equipment in the summary tables for measuring low-energy X-radiation, although this application is mentioned on page H-19.

g) General inconsistencies noted between table entries and the detailed descriptions should be corrected. These include differences in system names, systems missing from the tables in which they belong, and differences in cost estimates.

h) There is a lack of correspondence between the equipment mentioned in Chapters 6 and 7 with those in Appendix H. It is recommended that there be correspondence between the equipment mentioned in Chapters 6 and 7, the sheets in Appendix H, and Tables H.1 through H.5. Each sheet in Appendix H should be cross-referenced from the mentions of the equipment in the main body of the text.

4.4.2 Additional Measurement Techniques

Detailed descriptions should be added of the Frisch-Grid alpha detector for both field and laboratory use, and of the Pulsed-Laser Phosphorimetry (also known as Kinetic Phosphorescence Analysis, or KPA) for analysis of total uranium concentrations in the laboratory.

Appendix H should also address the advantages of isotopic analyses of some samples. In some cases, determination of isotopic ratios may be useful or even necessary to distinguish contamination from background, e.g., if the nuclide of concern is uranium and the contaminant is either depleted or enriched uranium with an isotopic signature readily distinguishable from that of natural uranium.

4.4.3 Relationship Between MARSSIM and MARLAP

The RAC was informed that ORIA is also currently involved in preparation of another multi-agency manual, the Multi-Agency Radiological Laboratory Analytical Protocols (MARLAP). In view of its expected overlap with MARSSIM's scope of coverage (especially Chapters 6, 7, and 9 and Appendix H), MARSSIM should discuss the relationship between MARSSIM and MARLAP and provide a short description of the scope of MARLAP in Chapter 1.

4.5 Quality Assurance and Quality Control

4.5.1 Data Quality Objective (DQO) Process as Related to Measurement Systems

Some danger exists that the Quality Assurance Program Plan (QAPP) will be too prescriptive and bureaucratic for a reasoned decision. The guidance given in MARSSIM Chapter 9 seems very specific and detailed, which may hinder development of reasonable site-specific plans. Hence, the document should encourage modification of the sample QAPP to fit site-specific conditions.

MARSSIM provides excellent guidance to showing how the Data Quality Objective (DQO) process is used to determine the number of measurements that must be performed—implicit in the selection of k_α and k_β probabilities (for detecting Type I and II errors). An equivalent level of detailed guidance should be provided for the establishment of specific criteria (e.g., bias and precision) for determining the quality or acceptability of analytical data that link the DCGL, measurement detection level, and analyte concentration or surface activity. This recommendation is consistent with guidance presented in MARSSIM Section 9.2.3, which states "The type and quality of environmental data needed for each project should be defined and documented using the DQO Process." For example, for cases in which the radionuclide of concern is not present in the background and in which the DCGL is one or two orders of magnitude higher than the analyte concentration and detection capability of the instruments, the data quality objectives in terms of bias and precision may be less stringent than when the analyte concentration is near the DCGL. In the latter case, much more restrictive bias and precision criteria would be required because a measurement error of a factor of two in analyte concentration could overlap the DCGL.

4.5.2 National Quality Assurance Standards

Industry guidance on the acceptable practices of a radioassay laboratory can be found in ANSI (1996). A description of several available government and industrial (interlaboratory) measurement assurance programs (MAPs) is presented in Appendix B of this ANSI standard. The ANSI standard should be mentioned during the discussion on the review of a laboratory's qualifications (MARSSIM, page 7-3, lines 77 through 114).

4.5.3 Data Verification and Validation

In the introductory material presented in Section 2, the term "Quality Assurance Project Plan (QAPP)" is used. More extensive discussion on the QAPP process is presented in Section 9. QAPP is an EPA term and concept dating to 1980 (U.S. EPA, 1980). Within the EPA QAPP guidance, detailed data verification and validation (Data V & V) specifications are delineated and required for data assessment. Although mentioned in Section 9.2.4 and Appendix E, MARSSIM does not discuss in any detail the Data V & V requirements or the extensive measurement documentation associated with the Data V & V process. (MARSSIM Chapter 9, line 114, refers the reader to Chapter 8, but there is no discussion of the topic in Chapter 8.) Furthermore, the glossary definitions of these terms are unintelligible for the average reader.

In the Introductory Section 9.1, it is stated that "The QAPP is a formal document describing in comprehensive detail the necessary QA, QC, and other technical activities that should be implemented to ensure that the results satisfy the stated objectives and to produce legally defensible data (EPA 1994c)." A writing group of ANSI was formed in 1995 to develop a consensus industry standard that addresses Radiation Data V & V. If MARSSIM is going to be a living document with routine updates, guidance from this ANSI standard can be referenced in the future.

Some of the Data V & V process has been incorporated into Section 9.4 but in a different format and intertwined throughout the six subsections of this section. Data V & V terminology and its data review cycle should be used explicitly and transparently in this section. MARSSIM should clarify whether the guidance it provides is sufficient to make the measurement documentation legally defensible in the absence of a formal Data V & V process.

4.5.4 Quality Control Samples and Measurements

Section 9.3, which deals with quality control (QC) samples and measurements, should be better organized to delineate between field QC measurements and samples, and laboratory QC samples. At present, due to the attempt to define and provide guidance on the types of QC samples, the reader may be confused as to the type and number of QC samples that are needed for field applications (or for an external laboratory QC program) compared to the requirements for an internal QC program at the laboratory.

Well-characterized performance evaluation samples are QC samples used by the field survey contractor to evaluate the performance of the laboratory to meet performance criteria for measurement quality as specified in the contractor's Statement of Work, e.g., detection limits, precision and bias specifications. These can be matrix spikes (including natural matrix spikes), duplicate samples (well-prepared split samples), or blanks (matrix sample without analyte). These samples are submitted as either double-or single-blind QC samples along with the field samples. Typically, the frequency with which performance evaluation samples are submitted is one per batch or one per twenty samples. The unique sample identifier for each performance evaluation sample is cross referenced or indexed with the samples in the same batch.

4.5.5 Use of Spikes

During the past five years, the operational QC programs for many national laboratory and uranium mining/milling site remediation projects have used site-specific natural matrix soil QC materials to ensure proper laboratory processing or gauge the quality of the very difficult radiochemical analyses. These natural matrix QC samples are used instead of matrix spikes (MARSSIM Section 9.3.2.1) wherein a soil matrix is spiked with an analyte as an ionic species in solution. In many cases the spiked matrix sample will be an ineffective QC material for monitoring certain laboratory radiochemical processes if the analyte is incorporated into the soil matrix of a survey sample rather than surficially adsorbed. Since the analyte in the natural matrix QC material is in the same chemical/physical form as the survey samples, these QC samples would be more effective in monitoring the laboratory's capabilities for the proper radioassay of the survey samples.

For natural matrix QC materials, a large amount of soil is collected from an area of known contamination and processed for the purpose of making laboratory performance evaluation (PE) material or samples. The contaminated soil material is homogenized by blending, or in some cases is dried, pulverized to micron size, and then blended. The blended soil is then split uniformly into many samples of a given weight. A statistically determined number of samples is selected to characterize the material for analyte concentration and other parameters. In some cases, several analyte concentration levels of the PE material can be prepared by blending a higher activity soil PE material with a blank soil material that has gone through the same preparation process and has the same particle-size distribution. The acceptable degree of sample analyte homogeneity and analyte distribution in the PE material for use in applying MARSSIM would be related to several factors: the DQOs of the project, the differential between the DCGL, measurement detection limit and expected residual analyte concentration at the survey site, and laboratory quality performance specifications. These PE samples are considered as true “natural matrix” spikes. The PE samples are included as double or single blind matrix spikes with each batch of samples submitted to the laboratory. This practice is consistent with the “Monitoring Laboratory” concept presented in ANSI (1996).

4.5.6 Data Quality Indicators

Under the current EPA Data V & V process, data qualifiers are assigned to each datum after the data package deliverables have been verified for compliance with the contractor’s Statement of Work and validated against qualitative and quantitative criteria. If MARSSIM suggests similar guidance, it is recommended that the Data V & V process and the final assignment of the data qualifiers be linked to reasonable data quality objectives that reflect the relationship between the DCGL, measurement detection limit, and the analyte concentration or surface activity (MARSSIM Section 9.4.6).

5. DEMONSTRATION OF COMPLIANCE

5.1 Response to Charge c)

The charge from ORIA's director regarding statistical and related issues was:

Charge c) Are the methods and assumptions for demonstrating compliance with a dose- or risk-based regulation technically acceptable? (b) Are the hypothesis and statistical tests and their method of application appropriate?

MARSSIM begins with the assumption that appropriate Derived Concentration Guideline Levels (DCGLs) have been defined that meet the dose- or risk-based release criteria for contamination of surface soil and building surfaces by radionuclides. It then proposes methods for dividing a site into survey units, surveying and sampling those units to develop distributions of radionuclide concentrations, and comparing the results to the DCGLs to decide whether the units pass the release criteria.

The Subcommittee made the following overall findings on this approach and the statistical issues involved in carrying it out:

a) Although the proposed approach is generally appropriate for demonstrating compliance with a dose- or risk-based regulation, it must be carefully applied to assure that it provides a technically acceptable solution to the *intent* of the regulations. While the designers of the MARSSIM are clearly aware of this issue and have prepared a document that *allows* the user to conduct an appropriate survey and analysis, more can be done to *assure* that they do so. In particular, MARSSIM allows the user to make appropriate choices for the hypothesis to be tested with the results of the survey and for the methods of conducting the statistical tests, but does not always provide sufficient guidance to make the best choice.

b) Although MARSSIM is applicable to the majority of contaminated sites, there appear to be cases that MARSSIM, as currently written, would have trouble addressing. These include: 1) cases dealing with the release of sites that had been contaminated with naturally occurring radionuclides and in which the DCGL is contained within the ambient (background) analyte variability, and 2) cases in which a reference background cannot be established. The Subcommittee recommends that future revisions of MARSSIM provide guidance to the user regarding appropriate choices when such conditions are encountered. For example, the null hypothesis might be redefined to be that the distribution of site radioactivity is no different from the reference site or than ambient radioactivity in general.

c) In MARSSIM, the preferred null hypothesis is that a survey unit is not ready for release and the information gathered must be sufficient, with a high degree of confidence, to accept the alternative hypothesis (i.e., that the unit meets the release criteria). Furthermore, MARSSIM discusses in detail two non-parametric procedures, the Wilcoxon Rank-Sum test and the Sign test, for testing this hypothesis. However, MARSSIM allows more flexibility in defining the null hypothesis and in choosing statistical analysis methods to test that hypothesis than may be readily apparent to most readers. This latitude needs to be more clearly stated and the criteria for selecting among potentially applicable tests need to be described.

In the sections that follow, the Subcommittee presents its detailed findings and recommendations regarding the above concerns. Additional comments have been transmitted under separate cover to ORIA.

5.2 Importance of Appropriate DCGLs

The MARSSIM team has decided that the best method of determining compliance with a dose- or risk-based standard is to define DCGLs—concentrations in soil or on surfaces that are unlikely to cause unacceptable doses and risks—and then test a survey area to see if those DCGLs are met. The Subcommittee notes that other approaches could have been proposed (e.g., using the measured concentrations to calculate dose or risk and then comparing the calculated value with the dose or risk criterion) but does not object to the method proposed. However, the Subcommittee believes that it is critically important for the assumptions and procedures used in MARSSIM to make comparisons with the DCGLs match those used in defining the DCGLs. For example, if a DCGL for soil is derived from a risk criterion by assuming that a receptor ranges over a certain area on a random basis, then the same area should be used for spatial averaging in the MARSSIM statistical analyses. Such averaging is usually performed from the standpoint of potential human receptors. The manuscript should note that different spatial and temporal scales of averaging will be necessary if dose- and risk-based criteria are applied to components of the ecosystem other than humans for derivation of a DCGL.

The three main steps in the MARSSIM process are: 1) determining the DCGLs from an agreed-upon release dose limit; 2) making accurate measurements of site contamination; and 3) determining whether the sampling data prove or disprove the hypothesis that the site contamination meets the release limits. (The Subcommittee recommends that in MARSSIM on page 1-3, lines 39-41, this process should be stated more neutrally as just done.) Nowhere, in the opinion of the Subcommittee, is the critical importance of the DCGLs given.

Furthermore, the document does not give any indication of how difficult the

DCGLs are to derive nor what the likely range of uncertainty would be under the most likely conditions. More prominent discussion should be given to the uncertainty associated with the derivation of the DCGLs. Often the DCGL is obtained using a computer model such as RESRAD (Yu et al., 1993) to back-calculate a value from a predefined risk or dose limit. There is considerable uncertainty associated with the results of this calculation. The effects of this uncertainty should be discussed. It is likely that the uncertainty associated with the mathematical derivation of a concentration guideline will greatly exceed variability of data from field samples. Information about the uncertainty in DCGLs is necessary so that the discussion of "Measure" and "Decide" would provide some idea of the accuracy needed in the derivation of DCGLs to be similar to that in the "Translate" portion of the study (MARSSIM, page 1-3, lines 33-41, and page 1-2, Figure 1.1). It would be useful to also give some idea of the techniques and models that can be used to determine the "Translate" portion of the study. Examples might include such models as MEPAS and RESRAD or the critique of the methodology by EPA or the NAS/NRC (Buck et al., 1995; Yu et al., 1993).

The use of any model to derive the DCGLs will induce uncertainty about how it relates to the dose or risk goal. Contrary to what is stated on MARSSIM page D-22, lines 514-527, this uncertainty can (and should) be quantified. Currently available models for characterizing sites with radioactive materials will not necessarily produce conservative estimates of risk unless conservative parameter choices are employed. DCGLs might best be derived in an iterative process, first using conservative screening models and parameter values and then, if the cost of sampling is prohibitively large, using site-specific models, parameter values, and uncertainty estimates to refine the uncertainty estimates for the DCGLs. Procedures for doing so can be found in NCRP Commentary No. 14 (NCRP, 1996).

The lack of specificity as to how DCGLs will be derived limits the reviewer's ability to comment on the propriety of the methods proposed in MARSSIM, because the assumptions and procedures used there must be consistent with those used in MARSSIM to ensure that the latter's methods are appropriate. For example, if the DCGL is based on uniform distribution in a semi-infinite slab, then measurement only in the top 15 cm may lead to incorrect release decisions (MARSSIM, page 1-4).

The issue of the derivation of the wide-area DCGL, $DCGL_w$, is critical. The $DCGL_w$ will change depending on the assumptions used to define the appropriate area for averaging of contamination to assess exposures to human and/or ecological receptors. The uncertainty in the derivation of the $DCGL_w$ may well exceed one order of magnitude. Consistency in the values of $DCGL_w$ is likely to change dramatically from site to site depending upon level of stakeholder involvement and the level to which stakeholders are risk-tolerant versus risk-averse.

The survey objectives seem to include the need to find radionuclides even in areas where they do not currently pose much risk (e.g., alpha-emitters under paint; beta-emitters behind walls; underground pockets of material). The DCGLs appear to be designed for materials that are readily available on wall surfaces and in surface soils. The possibility for a mismatch between the DCGL models and the MARSSIM survey methods seems significant. How would more appropriate scenarios for developing site-specific DCGLs be defined? (Pages 4-5, 4-21, 4-22).

If the maximum areas for the survey units are of the order of a large room or a medium-sized residential lot for building surface and soil contamination, respectively, are these survey areas consistent with assumptions underlying the calculations of DCGLs from the cleanup standard? For example, do the DCGLs always assume an infinite horizontal planar source? Would the DCGL be different if there were only one survey unit than it would be if there were several contiguous survey areas (MARSSIM page 4-13)

5.3 Relationship Between DCGL and Reasonable Measurement Requirements

There appear to be cases that MARSSIM, as currently written, would have trouble addressing. These cases include: a) cases dealing with the release of sites having been contaminated with naturally occurring radionuclides and in which the DCGL is nested in the ambient (background) analyte variability, and b) cases where no reference background for certain or old construction materials can be established. The Subcommittee recommends that future revisions of MARSSIM provide guidance to the user regarding appropriate choices when such conditions are encountered. For example, the null hypothesis might be redefined to be that the distribution of site radioactivity is no different from that at the reference site or than ambient radioactivity in general.

At the January 23, 1997 MARSSIM's SAB meeting, it was explained that for certain cases where MARSSIM may be difficult to apply, site surveys may be performed based on the "As Low As Reasonably Achievable (ALARA)" concept with concurrence of the regulator. The inference from such a philosophy is to set the Type I and Type II decision errors so that the probability of releasing the site is consistent with public policy goals. As stated in MARSSIM Section 2.5.4, "MARSSIM does not recommend values for any of these parameters (Type I or II decision errors, σ_s, σ_r or Δ), although some guidance is provided. A prospective power curve (see MARSSIM Appendix D) that considers the effects of these parameters can be very helpful in designing a survey and alternative values for these parameters, and is highly recommended." This ALARA concept should be presented in MARSSIM Sections 1 (page 1-3, line 43; page 1-4, line 67) and 2 (Section 2.5.4). If a survey plan using the ALARA approach is acceptable for such cases, then a specifically designed ALARA example of such an application should be included within the document.

5.4 Determination and Use of "Background" Measurements

MARSSIM provides good guidance on how to obtain background measurements from surveys of reference areas and how to use such measurements in the statistical analysis of results from a survey unit in order to determine its acceptability for release. How the reference areas are to be defined, located, and defended is less clear. The revised document should provide more guidance on the following issues:

a) If radionuclides are present in a survey unit as a result of human activities not related to the site in question, are they counted as background or site-related? (Pages 2-27, 4-11, 5-11)

b) MARSSIM does not make it clear whether background reference areas must always be on site, must always be off-site, or may be either. (Page 4-10)

c) It should be noted that representative reference areas cannot be found for many disturbed sites, such as former landfills. MARSSIM should provide at least a reference to identify acceptable characterization approaches for such sites.

d) MARSSIM should discuss whether subsurface background measurements should be taken at a "reference site". If so, MARSSIM, or the approved subsurface reference manual, should discuss how subsurface background measurements should be taken at a "reference site".

e) MARSSIM specifies that background reference areas must be similar to the survey areas with respect to radiological properties (MARSSIM pages 4-11 through 4-14). At the outset, this requirement seems to be an oxymoron. The document needs to provide more explanation regarding what is intended here.

f) Especially if a hostile party is on the decision team, agreement on what constitutes an acceptable background reference area may be difficult to achieve. For example, in the risk assessment performed for the cleanup of the Cotter uranium mill in Canon City, Colorado (ENVIRON Corp., 1991), definition of background was a key issue that limited the ability to make a widely shared decision. (Pages 4-2, 4-11)

MARSSIM defines "background radiation" to include fallout from nuclear weapons testing and nuclear power plant incidents in addition to radiation arising from natural sources. The Subcommittee notes that such a broad definition is widely—but not universally—accepted and the most common baseline used for evaluating levels of radioactive contamination. However, other guidance provided by various Federal agencies, including the EPA, restricts background radiation to include only natural sources, which may introduce some ambiguity for MARSSIM users (e.g., U.S. DOE, 1980; U.S. EPA, 1996; Mettler et al., 1990; U.S. Enrichment Corporation, 1993). The Subcommittee recommends that MARSSIM note the prevalence of other definitions of background, and that it discuss the scientific basis for its own broad definition. Wording similar to that used by the NRC in its recent decommissioning rule would be appropriate, in which the NRC defined background as follows (U.S. NRC, 1997):

> "Background radiation means radiation from cosmic sources; naturally occurring radioactive material, including radon (except as a decay product of source or special nuclear material); and global fallout as it exists in the environment from the testing of nuclear explosive devices or from past nuclear accidents such as Chernobyl that contribute to background radiation and are not under the control of the licensee. 'Background radiation' does not include radiation from source, byproduct, or special nuclear materials regulated by the Commission."

Defending this broad definition in its response to comments, the NRC stated that (NRC, 1997):

> "the Commission continues to believe that the inclusion in background of global fallout from weapons testing and accidents such as Chernobyl is appropriate. No compelling reason was presented that would indicate that remediation should include material over that the licensee has no control and that is present at comparable levels in the environment both on and offsite."

5.5 Statement of the Null Hypothesis and Statistical Tests

5.5.1 Statement of the Null Hypothesis

In MARSSIM, the preferred null hypothesis is that a survey unit is not ready for release. The information gathered must be sufficient, with a high degree of confidence, to accept the alternative hypothesis (i.e., that the unit meets the release criteria). The Subcommittee supports MARSSIM’s choice of this particular null hypothesis because it minimizes the potential for release of a survey unit for which insufficient information has been generated. Because the cost of surveying and sampling a site is usually—although not always—small in comparison to the costs of error in site classification, it is reasonable to adopt a strategy that encourages more rather than less investigation.

However, the Subcommittee can also support the use of a null hypothesis that presumes that a survey unit is no different from a reference unit where only background levels of radionuclides are present. With that hypothesis, the survey can be designed to reject the site for release only if the data show with a reasonable degree of confidence that residual radiological contamination at the site exceeds background levels plus the release criterion. Based on its conversations with the MARSSIM team, the Subcommittee believes that the document allows this type of hypothesis structure as a valid option if the site survey planning team agrees that it is a better approach for the specific situation. However, the document is less than explicit about this possibility, and it deserves more prominence, with appropriate cautions regarding its applicability.

5.5.2 Selecting Appropriate Statistical Tests

As with the selection of a hypothesis structure, MARSSIM also allows more flexibility in choosing statistical analysis methods than may be readily apparent to most readers. The document can easily be misinterpreted to confine the universe of permissible statistical tests to two non-parametric procedures, the Wilcoxon Rank-Sum test and the Sign test. Through more careful reading and conversations with the MARSSIM team, the Subcommittee was able to determine that other statistical tests could be more appropriate in specific situations and that the MARSSIM user was not restricted to the two specified tests. This latitude also needs to be more clearly stated, and the criteria for selecting among potentially applicable tests need to be described. Specifically, the non-parametric tests are designed to distinguish between distributions of concentrations that have different medians, and so are most applicable to reasonably symmetric distributions in which the median is likely to be a good approximation of the mean. This feature is important because the methods that relate dose or risk to concentration in the derivation of DCGLs use the average or mean concentration, not the median. Therefore, decidedly asymmetric distributions are probably better analyzed with methods other than the two featured in MARSSIM.

In the following sections, the Subcommittee details its comments and recommendations on the specified non-parametric tests, offers some suggestions regarding alternative methods, and expands on the issue of mean vs. median in statistical testing. Additional comments are also provided regarding three statistical design issues--the definition of the gray region, specification of sample size, and treatment of outliers.

5.5.3 Specified Non-Parametric Methods

While many users may not like their guidance to focus on statistics, MARSSIM may go a little too far in the direction of simplification. A reader may be led to believe that the Wilcoxon Rank Sum (WRS) test and the Sign test are all that one needs to do credible environmental statistics. These tests are useful tools, but hardly the last word in characterization statistics, and they will not perform well with markedly asymmetric (skewed) distributions.

In some cases of severely skewed distributions, the arithmetic mean can equal the 90th percentile of the distribution. In such a case, a sample of size 10 has about a 35% probability of having all observations below the true population mean. Thus one has an *a priori* 35% chance of declaring that a site whose mean slightly exceeds the DCGL is clean. Discussions with the MARSSIM team revealed the assumption that, for Class 1 areas, the need to detect areas of elevated concentration ("hot spots") would alleviate this problem. However, neither Table 5.8 nor Roadmap-9 states that hot spot detection must take place. If real hot spots exist, they will probably be detected on the preliminary scan, and the need to characterize these hot spots will increase the sample size, but how and when? The rationale for this assertion needs to be articulated better.

The Sign test presents further difficulties. Figure 1 shows a histogram of 1000 samples from a moderately skewed log normal distribution that, in the Subcommittee's experience, could easily be observed for environmental contamination. Its geometric mean (which is also the median in this case) and its geometric standard deviation are both equal to 2.72. Its arithmetic mean is about 4.5, approximately 1.6 times higher than the geometric mean and median and correspoonding roughly to the 70th percentile. If the DCGL were in the range 3 to 4, it would be lower than the arithmetic mean, and the site should not be released, but it would be higher than the geometric mean and median.

Suppose a sample of 11 observations is taken from this distribution. First, there is about a 2% (0.70^{11}) chance of drawing a sample where <u>all</u> observations are less than the mean. This rate is much lower than the 35% discussed above. However, the critical value for the Sign test is two or fewer values above the DCGL (probability $P \approx 0.03$). That is, if 2 or fewer of the observations exceed the DCGL, the site is considered "clean" (assuming no "hot" values). For our test distribution (which should fail meeting the release criterion), the probability of a "clean" result is about 0.3 because the probability of seeing a value above the mean is only about 30%. That is, our dirty site example would be released three times in ten, an unacceptable rate.

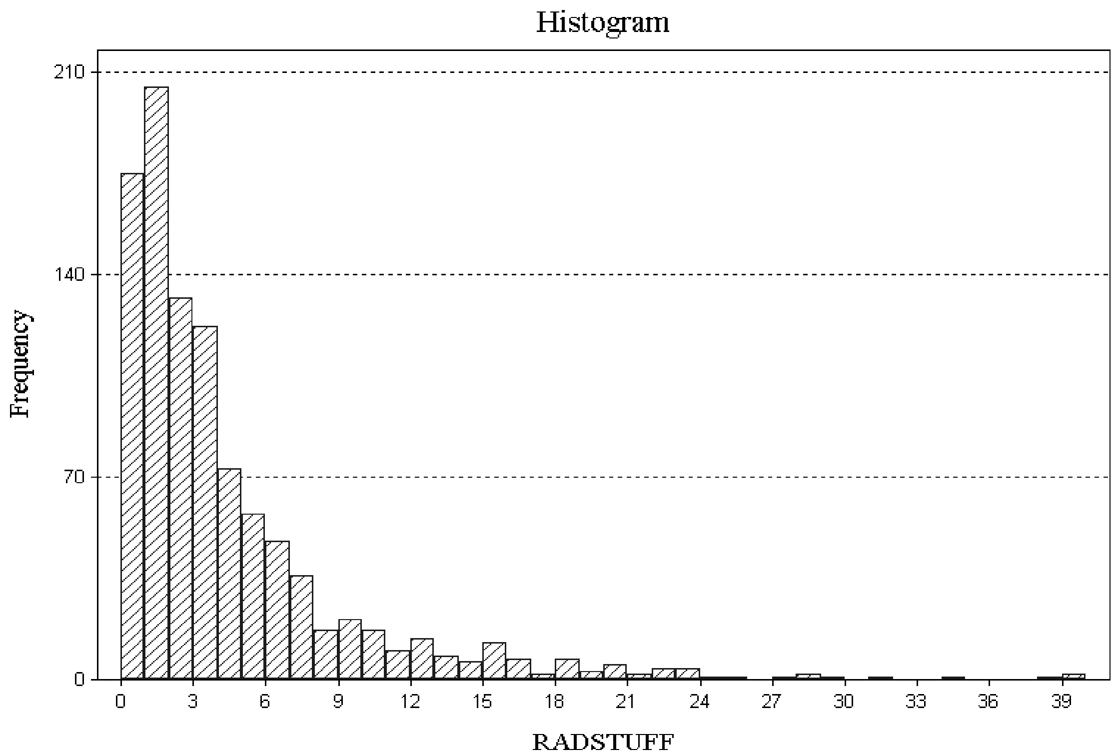

Figure 1. A hypothetical contamination distribution.

5.5.4 Importance of Distinguishing between Median and Mean

The target statistic for any exposure assessment should be the arithmetic mean concentration for a defined area, and the uncertainty associated with the estimate of the mean, due to all sources of potential error (variability of samples, analytical error, compromises in experimental design, and uncertainty due to differences in judgment amongst analysts). When the distribution of sample evidence is moderately to highly skewed, then non-parametric statistical techniques cannot be used to determine the uncertainty associated with the estimate of the arithmetic mean. These techniques are more appropriate for estimating uncertainty at specified quantiles (percentiles) of the sample distribution, including the 50th percent quantile (or median). As noted in the document, when the data are highly skewed, the median of the sample will underestimate the true arithmetic mean of surface contamination.

Model assumptions that use a uniformly distributed source term do so as a surrogate for the arithmetic mean of a heterogeneously distributed contaminant. If one hypothetically homogenized a heterogeneously contaminated area to produce a uniform contamination, the value of the uniform contamination would be equal to the arithmetic mean of the heterogeneously contaminated system.

The Sign test provides an indirect test of the mean only when the distribution of the sample is symmetrical in arithmetic space. In the experience of the Subcommittee, this is almost never the case. The majority of soil sampling programs usually reveals highly skewed distributions. Therefore, the Sign test, which is appropriate for testing differences in median concentrations, may not be appropriate to test for differences in mean concentrations.

The discussion about the mean and median on MARSSIM page D-9, lines 210-226, should be revised. For the purposes of limiting exposure as well as for the purposes of estimating exposure from a defined area, the target statistic should always be the arithmetic mean, regardless of whether the underlying distribution is symmetrical or skewed.

MARSSIM may allow the user to select a statistical test other than the WRS or Sign tests when a skewed, non-symmetric distribution is observed, but it does not make clear what is expected of the survey team when that occurs. The document should be explicit about whether other tests are to be considered at this juncture and how they should be selected. (Page 8-3, line 80; page 8-9)

5.5.5 Alternative Statistical Methods

The Subcommittee urges the MARSSIM team to consider, and perhaps encourage, alternative statistical methods for analyzing the survey data, especially when skewed distributions are encountered. Two techniques to be considered are discussed below: bootstrapping and Bayesian analysis.

Bootstrap estimator. The Sign test could be replaced by using a "resampling" or "— bootstrap" estimator for the distribution of the arithmetic mean (Efron and Tibshirani, 1993). Bootstrapping is a process that generates a series of estimates for the mean of a distribution by repeatedly resampling from the actual set of measured values, and then analyzes those means with standard statistical techniques. Such an approach is straightforward; we simply perform a large number K (e.g., K=1000) of iterations in which we resample, with replacement, from the original N sample values (sample values can occur more than once in each resampling), and calculate the mean of each iteration. For example, if N=10, the original sample might be 2,5,3,2,6,4,6,3,4,4, which has a mean of 3.9. The first resampling might yield 5,4,6,3,3,5,2,2,2,4 for a mean of 3.6. Additional resamplings would yield other means, both above and below 3.9. Depending on the skew of the original data and the number of iterations, the grand mean of the resampling might be higher or lower than 3.9, and we could also obtain an estimate of the uncertainty about that grand mean. The 50th largest mean value from our 1000 alternative realizations (of a sample of size N=10) would be equal to the upper 95th percentile of the true but unknown arithmetic mean. This 95th percentile value on the arithmetic mean must be less than the DCGL in order to declare the site safe for release. Using the bootstrap approach does not necessarily require one to generate a K of 1000 mean values; however, fewer mean values reduce the confidence of the final estimate of the mean. For example, if one had generated a K of only 45 mean values, then this data set K would produce a statistical tolerance limit that says that there is 90% confidence that there is at least a 95% chance that the true but unknown mean is less than that value. In this manner, the bootstrap method can be combined with the tools of non-parametric statistics to estimate uncertainty on the arithmetic mean of a sample (Conover, 1980). The foregoing is not to suggest that the non-parametric tests be abandoned, but rather that it might be a good idea for MARSSIM to discuss some possible alternatives.

Bayesian analysis. In some cases, the contamination data will not represent a truly random sample of the environment (e.g., data for hot spot samples). Such information can still be useful, but prior information about the sample's properties is needed, leading to a Bayesian view of hypothesis testing. When data are only partially representative of a remediated site, due to the fact that they are not taken from a randomized design or that they do not conform precisely to the same spatial and temporal scales as those upon which the DCGL is based, then classical statistical techniques are of limited use in determining the uncertainty about the true but unknown arithmetic mean concentration for that site. Under these circumstances, approaches based on Bayesian statistics may be advantageous (Carlin and Louis, 1996; Gelman et al., 1995). Bayesian statistics permit the explicit use of expert judgment to account for the inherent possibility of flaws and biases in the data. The result is that a credibility (or subjective confidence) interval can be obtained about the arithmetic mean (or any desired quantile) of the true but unknown distribution of soil concentration for both the remediated site and any reference site. These credibility intervals form the basis upon which subsequent decisions are made.

Nonparametric statistical tests are indeed superior to tests that are based on the assumption of an underlying normal distribution. However, the results of such nonparametric tests require that the samples be taken from a random or stratified random design. Nonparametric statistics cannot be used to estimate the uncertainty of the arithmetic mean of the sample. The target objective, however, should be a subjective confidence interval, within which the true but unknown arithmetic mean is, with a high degree of belief, contained. The proposed section in MARSSIM that discusses alternate statistical methods should include references to Bayesian approaches that will allow expert judgment to be combined with availability of imperfect sample evidence.

The use of professional judgment will be essential when data sets cannot be reliably evaluated using classical statistical procedures. Professional judgment is very useful in combination with Bayesian approaches so that the transition from imperfect sample evidence to subjective statements of confidence about the nature and extent of site contamination is transparent.

On Pg. 8-7 of MARSSIM, reference could be made to the use of Bayesian techniques to permit expert judgment to be used with sample evidence to determine the uncertainty of the mean value. If the variability among samples is very high and the number of samples are few (on the order of 6 to 10), it is possible for the maximum value to be less than the true mean for the system.

5.5.6 Definition of the Gray Region

The definitions of the terms, '*gray region*,' '*LBGR*' (Lower Bound of the Gray Region), and '*relative shift*' in MARSSIM are less clear than they should be and seem to differ among various parts of the document. It should be made clearer that the LBGR is a design choice that influences the sample size but does not enter directly into the release determination once the samples are taken. An analogy to the power specification in an epidemiological investigation might be useful. Moreover, it could be explained better that selecting the LBGR involves a tradeoff between spending too much on the survey and taking too large a chance that the site will fail even when the DCGL criterion is met. A small relative shift (LBGR close to DCGL) is appropriate when the decision is too close to call from *a priori* information. If the site is expected to pass easily, a lower LBGR (larger relative shift) may be appropriate. The sampling should be planned so that there is an acceptably small probability of rejecting a site that is, in fact, clean, given the cost of remediation.

It is not clear in MARSSIM how the gray area boundary conditions relate to the Type I and II decision errors, i.e., k_α and k_β probability set points. How would one relate the gray region boundary values to two sample distributions (reference background and actual analyte sample concentrations with associated standard deviations, σ_r and σ_s, respectively) separated by $4\sigma_r$, where an assumed equivalent standard deviation of $\sigma_r = \sigma_s$ (such as the distributions represented in MARSSIM Figure 6.2)? Guidance is also needed on whether or not there needs to be consistency of setting the Type I and Type II decision errors for different survey areas or survey type for the same site survey plan. For example, can the selection of the Type I and Type II decision errors (k_α and k_β probability values) be different for building surface areas, *in situ* field measurements, field sampling, core samples, etc.?

5.5.7 Specification of Sample Size

The Subcommittee agrees with the MARSSIM assumption that the desired bias should always be towards demanding an adequate number of data points. The cost of additional sampling and analysis will usually be small with respect to the cost of cleanup. In those cases where the cost of cleanup is small with respect to the cost of sample analysis, then it may be cost effective to clean up using conservative estimates of a DCGL to ensure that contamination has been effectively removed.

If we want to be sure of sampling the ω upper percentile (say 95th) of the distribution with some large probability P (e.g., P=0.95), the relevant equation for the required sample size N is:

$$N = \ln(1-P) / \ln(\omega) \qquad [1]$$

The result of [1] with P and ω = 0.95 is 59 if we round up to the nearest integer. Therefore, a sample size of about 60 will nearly always be sufficient to characterize a survey unit.

5.5.8 Treatment of Outliers

It is inevitable that some investigations will yield one or more results that seem inconsistent with the remaining data and other information. There is no discussion in MARSSIM of statistical tests for identifying these “outliers.” Although they cannot be rejected *a priori*, they should be viewed with caution, especially when they are sufficiently high to shift the test result from “pass” to “fail.” The MARSSIM document is not as clear as it should be regarding the treatment of outliers. Although it guides the user toward further investigation when a value exceeds an “investigation level,” it is not explicit about what investigations are appropriate or how any resampling results should be combined with the original value. Whether or not the value of the measurement exceeds an investigation level, what responses are appropriate for an outlier, especially if it is well above the typical range? How should outliers be treated when there is no obvious failure in application of quality assurance or quality control practices? How are remeasurements weighted when they differ from the original measurements? MARSSIM should provide guidance to the user as to how to identify and treat outliers, or a justification for including such data in subsequent statistical calculations.

6. BROADER ISSUES

The following comments, while not directly in response to the charge, may help to improve the MARSSIM document and its use. Included are comments on application of MARSSIM outside site boundaries, the conservatism of MARSSIM, and the future of MARSSIM after issuance of the final document.

6.1 Application of MARSSIM Outside Site Boundaries

Many considerations enter the decision whether or not a site can or should be released. One issue is that of the radiological condition of off-site areas, such as those in which contamination may have been introduced in the past (or during remedial activities) by surface runoff or windblown material. MARSSIM does not make it clear how areas not within the site boundaries but potentially impacted by site activities would be treated. When are they subject to survey? Are off-site survey areas subject to the same rules as on-site ones? While such areas may need cleanup to reach risk- or dose-based criteria, they may already be unrestricted in use, so “release” is not, strictly speaking, an issue. However, access to potentially impacted lands outside site boundaries is often a significant problem.

The Subcommittee recognizes that application of the MARSSIM approach to investigating radioactive contamination in off-site areas is a policy decision to be made by the governing Federal agency. However, it would be prudent for MARSSIM to urge the user to be vigilant during the Historical Site Assessment in identifying "vicinity properties" that may be contaminated as a result of contamination from, and waste disposal practices at, the primary or secondary sites. Ancillary sites that may need to be checked include landfills, sewage treatment plants (STP), STP sludge disposal sites, dredge spoils disposal sites, site operations spoils disposal sites, and sewer maintenance disposal sites.

6.2 Conservatism of MARSSIM

If default DCGLs are derived using risk assessment methods similar to those used in the EPA Draft Technical Support Document for the Development of Radionuclide Cleanup Levels for Soil (U.S. EPA, 1994), they will be conservative in the sense that they will more often than not be lower than necessary to achieve the risk or dose goals for site release. MARSSIM also appears to be conservative, in the sense that it seems more likely to reject a site that meets the DCGLs than to release a site that does not. Its principal conservatism lies in its selection of the recommended null hypothesis: "The residual radioactivity in the survey unit exceeds the release criterion." The degree to which that choice will bias release decisions will depend on the values specified for Type I and Type II decision errors (α and β, respectively).

Therefore, MARSSIM has the potential to compound the conservatism likely to be present in the DCGLs if they are not made site-specific. Although this extra margin of safety may be prudent public policy, the decision-makers ultimately responsible for site release need to understand its existence and, to the extent possible, its magnitude. Furthermore, any conservatism in translating a dose or risk criterion into DCGLs and in determining compliance with those DCGLs through MARSSIM was probably not included in the cost/benefit analysis for the cleanup standard, which could result in lower cleanup benefits or greater cleanup costs than expected by those who will eventually set the cleanup standards.

Although not calling for a quantitative analysis of the effects of compounded conservatism, the Subcommittee recommends that MARSSIM include a qualitative summary of its assumptions and policy choices, showing the likely direction of any biases introduced and, if possible, the relative magnitude of such biases. It should also recommend that the planning team reveal its own assumptions and choices when a site-specific survey design is developed.

6.3 Post MARSSIM

A process should be established before MARSSIM is finalized to provide for future revisions that will reflect changes in regulations and agency policies as well as improvements as experience is gained in applying MARSSIM to real sites. For example, once MARSSIM is complete, will each of the federal agencies make subsequent revisions to it on their own, or will MARSSIM continue as a multi-agency collaborative effort in the future? Each revision should be reflected in the NUREG, EPA, and NTIS publication numbers. Clearly, it is preferable to keep MARSSIM a multi-agency document after it is finalized.

As needs arise, this successful interagency approach to survey surficial contamination could be applied to radiological surveys of other media, such as groundwater, subsurface soils and sewer contamination. In addition, the interagency approach could be applied to activities such as site stabilization, decommissioning techniques, and standardized sampling procedures for various media.

7. FINDINGS AND RECOMMENDATIONS

7.1 Overall Approach to Planning Surveys

7.1.1 ***Value of MARSSIM***. The MARSSIM document brings together DOE, EPA, NRC, and DOD with a common method of site surveys and investigations. Adopting MARSSIM will mean that surveys done for any of the agencies will be immediately transparent to all. In general, the Subcommittee found that MARSSIM is nearly a finished product. The multi-agency team is commended for its work in addressing the many complex issues involved, resulting in the compilation of an exceptionally well-prepared reference which is technically sound and which will be a useful tool for guiding final status surveys. The document provides generally consistent and explicit guidance for planning and conducting radiation surveys for the decommissioning of radiologically contaminated sites. A major value of MARSSIM is providing managers with the quantitative tools necessary for determining how much to budget for sampling and analytical efforts.

7.1.2 *Use of references***.**

The credibility of the MARSSIM report would be greatly enhanced by an improved reference section that includes recognized, readily accessible, up-to-date literature.

7.1.3 *Consistency with federal regulations***.**

MARSSIM has the challenge and responsibility to provide information that is consistent with the current regulatory picture. Because this aspect is admittedly somewhat of a moving target, the Subcommittee recommends that MARSSIM avoid referencing regulations presently in draft form and instead refer to the regulatory responsibilities of the agencies in these cases.

7.1.4 *Additional considerations for site release***.**

It needs to be emphasized early in MARSSIM that compliance with the DCGL is only one of the considerations for release of a site for unrestricted use. Other considerations not addressed in MARSSIM include the radioactive contamination of subsurface soil, surface water, or ground water. Furthermore, if the release criterion is supposed to be ultimately risk-based, risks of residual chemicals may also need to be factored into the decision about release criteria for radionuclides.

7.1.5 *Scope limited to contaminated surfaces*.

MARSSIM should discuss its rationale for limiting its scope to guidance for contaminated surface soils and building surfaces. Furthermore, it should more clearly state that radioactive contamination of subsurface soil, surface water, and ground water are explicitly excluded from its coverage. The document should include some discussion of why these particular media were not included and the plans, if any, to cover them in the future. The contents of MARSSIM should then be made consistent with its stated scope of coverage throughout, such that sections not clearly falling within the scope should be omitted or moved to an appendix. If these sections are left in the report in any form, then they raise the question as to whether or not the DCGLs established for surface soils and building surfaces also apply to these media. Also, MARSSIM should discuss the extent to which it is necessary to evaluate scenarios under which subsurface contamination might be expected to contribute to surface contamination in the future, and how this affects the decision of whether the site meets release criteria. For example, contaminated subsurface soil could be exposed via erosion or excavation, and contaminated groundwater could result in surface contamination via seepage and direct pumping of the groundwater.

7.1.6 *Consistency between sampling methods and assumptions underlying derivation of DCGL*.

It is critical to verify the consistency between the definition of surface soil and the method by which the DCGL is established from pathway modeling. For example, it is not entirely clear whether a DCGL applies to soil at the surface, at all depths, or only to a certain depth (e.g., 15 cm). Conversely, it is critical that the user ensure that the manner in which samples are collected and analyzed is consistent with the underlying assumptions about distribution of radioactive contaminants on which the DCGL is based.

7.1.7 *Public involvement*.

The MARSSIM document is not very clear about the composition or functioning of the planning team, especially the issue of public participation. This issue should be raised in the document and, to the extent possible, activities in the agencies regarding public involvement in site survey design and implementation should be referenced there.

7.2 Data Acquisition and Assessment

7.2.1 *Organization of information concerning data acquisition*.

Sample collection protocols and analytical techniques are discussed in MARSSIM Chapters 6 and 7 and in Appendix H. The redundancy in coverage between Chapters 6 and 7 should be minimized by careful editing and cross-referencing, by redefining the scope of each chapter, or by combining the two chapters into a single one, resulting in a more concise, user-friendly, and internally consistent document. Standardized nomenclature should be used throughout MARSSIM in references to field and laboratory equipment in order to avoid confusion or ambiguity. All techniques mentioned in the main body of MARSSIM should have corresponding detailed descriptions in Appendix H.

7.2.2 *Field Measurement Methods*.

Descriptions of field measurement methods, instruments, and operating procedures in MARSSIM are technically sound but incomplete. Some additions, clarifications, and corrections are noted in our report. MARSSIM should provide guidance for the development of standardized procedures including a list of considerations for designing site-specific surface-soil sampling and preparation methods so as to ensure that samples will be representative of the materials of concern in deriving the DCGLs for the site.

7.2.3 *Field and laboratory instrumentation*.

Descriptions of the selection and operation of radiation detection instruments for laboratory analyses are technically sound and represent standard practice but may not be state-of-the-art. MARSSIM should standardize the level of detail used in its presentation of this material and should also provide information on the planned scope and current status of plans to prepare a manual on Multi-Agency Radiological Laboratory Analytical Protocols (MARLAP), which may be a more appropriate forum in which to provide more thorough in-depth guidance to the user on the selection and operation of laboratory instrumentation.

7.2.4 *Sample collection and preparation for analysis*.

Surface-soil sample collection and preparation methods described in the manual are technically sound but incomplete. MARSSIM should provide guidance for the development of standardized methods, referencing relevant ASTM standards. MARSSIM should also provide a list of additional considerations for designing site-specific surface-soil sampling and preparation methods so as to ensure that samples will be representative of the materials of concern in deriving the DCGLs for the site.

7.3 Demonstrating Compliance

7.3.1 ***Consistency between sampling data and the basis for definition of DCGLs***.

The Subcommittee believes that it is critically important that the assumptions and procedures used in MARSSIM to make comparisons with the DCGLs match those used in defining the DCGLs. For example, if a DCGL for soil is derived from a dose limit or risk criterion by assuming that a receptor ranges over a certain area on a random basis, then the same area should be used for spatial averaging in the MARSSIM statistical analyses. Such averaging is usually performed from the standpoint of potential human receptors. The manual should note that different spatial and temporal scales of averaging will be necessary if dose and risk based criteria are applied to components of the ecosystem other than humans for derivation of a DCGL. This recommendation assumes that the DCGL is derived in a manner appropriate for characterizing human and/or ecological exposures likely to occur at the site under investigation.

7.3.2 ***Cases for which the MARSSIM approach may not be able to demonstrate compliance***.

There appear to be cases that MARSSIM, as currently written, would have trouble addressing. These cases include: 1) cases dealing with the release of sites having been contaminated with naturally occurring radionuclides and in which the DCGL is contained within the ambient (background) analyte variability, and 2) cases where no reference background for certain or old construction materials can be established. The Subcommittee recommends that future revisions of MARSSIM provide guidance to the user regarding appropriate choices when such conditions are encountered. For example, the null hypothesis might be redefined to be that the distribution of site radioactivity is no different from that at the reference site or than ambient radioactivity in general.

7.3.3 ***Uncertainty in DCGL values.***

MARSSIM properly warns the user that the DCGL is not free of error and that the uncertainty associated with this quantity may be considerable if derived using generic assumptions and parameter values. However, its discussion of this issue is relegated to an appendix. This important aspect, with an expanded discussion of its implications for the release decision, need to be disclosed more prominently in the text of the main document. It is clearly undesirable to design a survey around a DCGL that may not be relevant to the actual conditions at a site, such that actual exposures, doses, and risks would be largely different than those used to derive the generic DCGL. Consequently, MARSSIM should more strongly encourage the user to examine critically the assumptions made in any model used to derive DCGLs for a site in order to determine whether application of site-specific information and parameters would result in large modifications to the proposed DCGL, or whether development of a site-specific model would be warranted in order to obtain a DCGL that is more relevant to the human and ecological exposure conditions prevailing at the site.

7.3.4 ***Reference areas for background measurements*.**

MARSSIM provides good guidance on how to obtain background measurements from surveys of reference areas and how to use such measurements in the statistical analysis of results from a survey unit in order to determine its acceptability for release. How the reference areas are to be defined, located, and defended is less clear. For example, if radionuclides are present in a survey unit as a result of human activities not related to the site in question, are they counted as background or site-related? Must background reference areas always be on- site, off-site, or either? What should be done if a representative reference area does not exist, e.g., for a disturbed site such as a former landfill?

7.3.5 ***Flexibility in statistical procedures*.**

In MARSSIM, the preferred null hypothesis is that a survey unit is not ready for release, and the information gathered must be sufficient, with a high degree of confidence, to accept the alternative hypothesis (i.e., that the unit meets the release criteria). Furthermore, MARSSIM discusses in detail two non-parametric procedures, the Wilcoxon Rank-Sum test and the Sign test, for testing this hypothesis. However, MARSSIM allows more flexibility in defining the null hypothesis and in choosing statistical analysis methods to test that hypothesis than may be readily apparent to most readers. This latitude needs to be more clearly stated and the criteria for selecting among potentially applicable tests need to be described.

7.3.6 ***Distinction between median and mean*.**

MARSSIM’s discussion about the mean and median should be revised in order to ensure that the correct statistical parameter is used to compare concentrations in the survey area to those in the reference area. The target statistic for any exposure assessment should be the arithmetic mean concentration for a defined area, together with the uncertainty associated with the estimate of the mean. For a normally distributed population, the mean and median are identical in values. However, when the distribution of sample evidence is moderately to highly skewed, then non-parametric statistical techniques cannot be used to determine the uncertainty associated with the estimate of the arithmetic mean, and the median of such a sample set will underestimate the true arithmetic mean of surface contamination. The majority of soil sampling programs usually reveal highly skewed distributions. Therefore, the Sign test, which is appropriate for testing differences in median concentrations, may not be appropriate to test for differences in mean concentrations.

7.3.7 *Alternative statistical methods*.

The Subcommittee urges the MARSSIM team to consider, and perhaps encourage, alternative statistical methods for analyzing the survey data, especially when skewed distributions are encountered. Two techniques suggested for consideration are bootstrapping and Bayesian analysis.

7.3.8 *Treatment of outliers*.

It is inevitable that some investigations will yield one or more results that seem inconsistent with the remaining data and other information. MARSSIM does not discuss statistical tests for identifying these "outliers." Although they cannot be rejected *a priori*, they should be viewed with caution, especially when they are sufficiently high to shift the test result from "pass" to "fail."

7.4 Broader Issues

7.4.1 *Conservatism of MARSSIM*.

The guidance provided by MARSSIM may introduce an additional measure of conservatism in the process of setting and determining compliance with radiation cleanup standards, compounding the conservatism already likely to occur in developing default DCGLs. Release decisions may be biased correspondingly. MARSSIM should include a qualitative summary of any biases that may result from its assumptions and policy choices, and recommend that the planning team be similarly revealing when developing a site-specific survey design.

7.4.2 *Future revisions of MARSSIM*.

Before MARSSIM is finalized, a process should be established to provide for future revisions that will reflect changes in regulations and agency policies as well as improvements as experience is gained in applying MARSSIM to real sites.

7.4.3 *Extension of MARSSIM and development of other multi-agency guidance manuals*.

As needs arise, this successful interagency approach to survey surficial contamination could be applied to radiological surveys of other media, such as groundwater, subsurface soils and sewer contamination. In addition, the interagency approach could be applied to activities such as site stabilization, decommissioning techniques, and standardized sampling procedures for various media.

7.4.4 *Evaluation of methods for deriving DCGLs*.

DCGLs are critical for determining the acceptability of residual levels of radioactivity remaining after a site has been remediated. The Subcommittee suggests that the various approaches proposed for derivation of DCGLs be reviewed and evaluated. This evaluation can be performed by an interagency group and by the EPA/SAB. This evaluation should focus on the strengths and weaknesses of current methodologies and opportunities to refine generic DCGLs with improved site-specific models and data. This review is important but outside the scope of MARSSIM *per se.*

APPENDIX A- REFERENCES

American National Standards Institute (ANSI), 1996. “Measurement and Associated Instrument Quality Assurance for Radioassay Laboratories,” ANSI N42.23 - 1996.

American Society of Testing and Materials (ASTM), 1983. ASTM C998-83, “Standard Method for Sampling Surface Soil for Radionuclides.”

American Society of Testing and Materials (ASTM), 1983. ASTM C999-83, “Standard Method for Soil Sample Preparation for the Determination of Radionuclides.”

Buck, J.W., G. Whelan, J.G. Droppo, D. L. Strenge, K.J. Castleton, J.P. McDonald, C. Sato, and G.P. Streile, 1995. Multimedia Environmental Pollutant Assessment System (MEPAS) Application Guidance, Guidelines for Evaluating MEPAS Input Parameters for Version 3.1, Report PNL-10395, Pacific Northwest Laboratory (Richland, Washington).

Carlin, B.P., and Louis, T.A. 1996. Bayes and Empirical Bayes Methods for Data Analysis. Monographs on Statistics and Applied Probability 69. Chapman & Hall, St. Edmundsbury Press, Bury St. Edmunds.

Conover, W. J., 1980. Practical nonparameteric statistics. Wiley & Sons, NY.

Currie, L.A., 1968. Limits for Qualitative Detection and Quantitative Determination. *Analytical Chemistry*, 40(3):586-693.

Efron, B., and R.J. Tibshirani, 1993. An Introduction to the Bootstrap. Chapman and Hall, New York.

Eisenhower, E.H., 1991. Criteria for the Operation of Federally-Owned Secondary Calibration Laboratories, National Institute of Standards and Technology (NIST) Special Publication 812, October 1991.

ENVIRON Corporation, 1991. Health Risk Assessment of the Cotter Uranium Mill Site, Caňon City, Colorado. October 29.

Federal Register, Vol. 62, No. 139, Monday, July 21, 1997, p. 39087 (Reference on Background Radiation Definition.)

Gelman, A., Carlin, J.B., Stern, H.S., and Rubin, D.B. 1995. Bayesian Data Analysis. Chapman & Hall, St. Edmundsbury Press Ltd., Bury St. Edmunds.

Health Physics Society (HPS), no date. Calibration Laboratory Accreditation Program, Criteria for Accreditation of Calibration Laboratories, Laboratory Accreditation Assessment Committee of HPS.

Kooyoomjian, J., 1997. Memorandum to L.G. Weinstock, M.E. Clark, M.P. Doehnert, B. Littleton, and C. Petullo concerning "Transmittal of Detailed Comments on the Multi-Agency Radiation Survey and Site Investigation Manual (MARSSIM)," dated July 24, 1997.

Mettler, F., et al., 1990. Medical Management of Radiation Accidents. CRC Press. Definition of "background" on page 374 of the Glossary.

NCRP (National Council on Radiation Protection and Measurements), 1996. A Guide for Uncertainty Analysis in Dose and Risk Assessments Related to Environmental Contamination. NCRP Commentary No. 14. NCRP, Bethesda, MD.

U.S. DOE (U.S. Department of Energy), 1980. A Background Report for Formerly Utilized Manhattan Engineer District/Atomic Energy Commission Sites Program. Report DOE/EV-0097A. September 1980. Definition of "background radiation" on page 173 of the Glossary.

U.S. Enrichment Corporation, 1993. Environmental Assessment for the Purchase of Russian Low Enriched Uranium Derived from Dismantlement of Nuclear Weapons in the Countries of the Former Soviet Union. Draft dated November 25, 1993. Definition of "background radiation" on page 10-1 of the Glossary.

U.S. EPA (U.S. Environmental Protection Agency), 1980. Interim Guidelines and Specifications for Preparing Quality Assurance Program Plans. QAMS-005/80, EPA, Washington D.C.

U.S. EPA (U.S. Environmental Protection Agency), 1994. Technical Support Document for the Development of Radionuclide Cleanup Levels for Soil. Review draft dated September 1994.

U.S. EPA (U.S. Environmental Protection Agency), 1996. Technology Screening Guide for Radioactively Contaminated Sites. Report prepared under Contract No. 68-D2-0156 for the Office of Radiation and Indoor Air (ORIA) and the Office of Solid Waste and Emergency Response (OSWER). Report EPA 402-R-96-017. November 1996. Definition of "background radiation" on page A-6.

U.S. EPA (U.S. Environmental Protection Agency), U.S. Department of Energy (DOE), U.S. Department of Defense (DOD), and U.S. Nuclear Regulatory Commission (NRC), 1996. Multi-Agency Radiation Survey and Site Investigation Manual (MARSSIM). Draft for public comment, December 6, 1996. EPA 402-R-96-018, NUREG-1575, NTIS-PB97-117659.

U.S. NRC (U.S. Nuclear Regulatory Commission), 1997. NRC Final Rule on Radiological Criteria for License Termination, 10 CFR Part 20 and Conforming Amendments in 10 CFR Parts 30, 40, 50 and 51, 62 Federal Register (FR), 39058, July 21, 1997.

Yu, C., A.J. Zielen, J.-J. Cheng, Y.C. Yuan, L.G. Jones, D.J. LePoire, Y.Y. Wang, C.O. Loureiro, E. Gnanapragasam, E. Faillace, A. Wallo III, W.A. Williams, and H. Peterson, 1993. Manual for Implementing Residual Radioactive Material Guidelines Using RESRAD Version 5.0. ANL/EAD/LD-2, Argonne National Laboratory.

APPENDIX B - LIST OF ACRONYMS

α	Type I decision error
β	Type II decision error
Δ	relative shift
σ_r	one-sigma uncertainty of the mean background concentration
σ_s	one-sigma uncertainty of the mean sample concentration
ALARA	As Low as Reasonably Achievable
ANSI	American National Standards Institute
ARAR	Applicable and Relevant or Appropriate Requirements
ASTM	American Society of Testing Materials
Bq	Bequerel (unit of radioactivity)
Ci	Curie (unit of radioactivity)
Cd	cadmium
CERCLA	Comprehensive Environmental Response, Compensation, and Liability Act
CFR	Code of Federal Regulations
cm	centimeter
Co-60	cobalt-60, a radioactive isotope of cobalt
DCGL	Derived Concentration Guideline Level
$DCGL_{EMC}$	Elevated Measurement Comparison DCGL
$DCGL_w$	Wide-area DCGL
DOD	Department of Defense
DOE	Department of Energy
dpm	decays per minute
DQO	Data Quality Objective
EPA	Environmental Protection Agency (U.S. EPA)
h	hour
H-3	hydrogen-3 (tritium), a radioactive isotope of hydrogen
k_α	probability of a Type I decision error
k_β	probability of a Type II decision error
kg	kilogram
LBGR	Lower Bound of Gray Region
m	meter
MAP	measurement assurance program
MARLAP	Multi-Agency Radiation Laboratory Analytical Procedures (Manual)
MARSSIM	Multi-Agency Radiation Survey and Site Investigation Manual
MDC	Minimum Detectable Concentration
mSv	milli-Sievert
NaI	sodium iodide
NAS	National Academy of Science
NCRP	National Council on Radiation Protection and Measurements

Ni-63	nickel-63, a radioactive isotope of nickel
NIST	National Institute of Standards and Technology
NRC	Nuclear Regulatory Commission
NTIS	National Technical Information Service document
NUREG	Nuclear Regulatory Commission document
ORIA	Office of Radiation and Indoor Air
PE	performance evaluation
PERALS	Photon Electron Rejecting Alpha Liquid Scintillator
PIC	Pressurized Ionization Chamber
QA	quality assurance
QAPP	Quality Assurance Project Plan
QC	quality control
R	Roentgen
RAC	Radiation Advisory Committee (U.S. EPA/SAB/RAC)
RCRA	Resource Conservation and Recovery Act
RESRAD	Residual Radioactive Material (transport code, which includes dosimetry and other components)
RI/FS	Remedial Investigation/Feasibility Study
SAB	Science Advisory Board (U.S. EPA/SAB)
SI	Le Systeme Internationale de Unites (International System of Units)
SOW	statement of work
STP	sewage treatment plant
Te	tellurium
UMTRCA	Uranium Mill Tailings Radiation Control Act
V&V	verification and validation
WRS	Wilcoxon Rank Sum (statistical test)
Zn	zinc

APPENDIX C - GLOSSARY

accuracy - Closeness to the true value; the extent to which a given measurement agrees with the standard value for that measurement: the degree of agreement of the observed value with the true value of the quantity being measured.

action level - The numerical value that causes the decision maker to choose one of the alternative actions. It may be a regulatory threshold standard (e.g., Maximum Contaminant Level for drinking water); a dose- or risk-based concentration (e.g., DCGL); or a reference-based standard.

ALARA (As Low as Reasonably Achievable) - A basic concept of radiation protection which specifies that exposure to ionizing radiation and releases of radioactive materials should be managed to reduce collective doses as far below regulatory limits as is reasonably achievable considering economic, technological, and societal factors, among others. Reducing exposure at a site to ALARA strikes a balance between what is possible through additional planning and management, remediation, and the use of additional resources to achieve a lower collective dose level. A determination, and the use of additional resources to achieve a lower collective dose level. A determination of ALARA is a site-specific analysis that is open to interpretation, because it depends on approaches or circumstances that may differ between regulatory agencies. An ALARA recommendation should not be interpreted as a set limit or level, and this system of dose limitation which is based on keeping exposures "as low as reasonably achievable," also takes economic and social factors into account.

alpha particle - Two neutrons and two protons bound as a single particle that is emitted from the nucleus of certain radioactive isotopes in the process of decay or disintegration.

alpha-scanning equations - Equations used to determine the probability of detecting a particular level of surface contamination using an alpha survey meter or to calculate a scan rate which will assure that a particular level of contamination can be detected with a specific level of confidence.

ARAR (Applicable and Relevant or Appropriate Requirements) - A guidance value (often a concentration level) taken from another regulatory context (such as maximum contaminant levels for drinking water), but applied as a criterion for cleanup of hazardous waste [e.g., Superfund, Resource Cconservation and Recovery Act (RCRA) and other] sites.

area - A general term that refers to any portion of a site, up to and including the entire site.

arithmetic mean - The average value obtained when the sum of individual values is divided by the number of values.

background (instrumental, method or process blank, field blank, environmental) - The normal level observed, above which the phenomenon under investigation must be detected.

background radiation - Ambient signal response recorded by measurement instruments that is independent of radioactivity contributed by the radionuclides being measured in the person or sample. It includes radiation from cosmic sources; naturally occurring radioactive material, including radon (except as a decay product of source or special nuclear material); and global fallout as it exists in the environment from the testing of nuclear explosive devices or from nuclear accidents like Chernobyl which contribute to background radiation and are not under the control of the cognizant organization. Background radiation does not include radiation from source, byproduct, or special nuclear materials regulated by the cognizant Federal or State agency.

Bayesian analysis - An approach to data analysis that uses both the information contained in a data set and prior information in the form of a probability distribution, concerning the likely value of the quantity of interest. For example, a Bayesian estimate of the sample mean would use both the sample data, and a prior distribution for the value of the sample mean. The prior distributions can be based on historical data or on a "personal" probability distribution assumed by the data analyst.

According to the Bayesian view, all quantities are of two kinds; those known to the person making the inference and those unknown to the person. The former are described by their known values, the uncertainty surrounding the latter being described by a joint probability function for them all. Bayesian statistics permit the explicit use of expert judgement to account for the inherent possibility of flaws and biases in the data. The result is that a credibility (or subjective confidence) interval can be obtained about the arithmetic mean (or any desired quartile) of the true, but known, distribution and any reference site.

In an analysis in which Bayes' theorem is used to derive posterior probabilities from assumed prior knowledge together with observational data, for example, biological information on the relationship between species and hazardous substances can be combined with data on interspecies dose response to calculate the response of human populations.

bias - A type of measurement error where each observation is mismeasured by a constant amount. For example, if we measure objects with a "12 inch" ruler that is really 11 inches long, all measurements will be too short by about 8.4%. This is a systematic or persistent distortion of a measurement process which causes errors in one direction; the systematic or persistent distortion of a statistic may be as a result of sampling procedure or other anomaly.

In ANSI N13.30-1989D, bias is defined as “(a) The deviation of the expected value of a random variable from a corresponding correct value. (b) A fixed deviation from the true value that remains constant over replicated measurements within the statistical precision of the measurement. (Synonyms: deterministic error, fixed error, systematic error.)

bootstrap estimator - An estimate of the variability of some statistic of interest such as the sample mean, which is obtained by repeatedly sampling with replacement, a set of measured values. For example, if one has a set of N observations and is interested in the variability of the sample mean, one would take, say, 1000 resamples of N from the sample data (because we sample with replacement, some values will occur more than once in a given sample) and examine the distribution of the means of these resamples. Thus, the bootstrap estimator is an estimate of some statistic of a distribution (such as its mean) derived by repeatedly sampling from a set of measured values from that distribution.

characterization survey - A type of survey that includes facility or site sampling, monitoring, and analysis activities to determine the extent and nature of contamination and to provide the basis for acquiring necessary technical information to develop, analyze, and select appropriate cleanup techniques.

cleanup - Actions taken to prevent, minimize, or mitigate damage to the public health or welfare or to the environment, which may otherwise result from a release or threat of release of a hazardous substance to the environment. Cleanup is sometimes used interchangeably with the terms remedial action, response action, or corrective action.

coincidence counting - Recording the occurrence of counts in two or more detectors simultaneously or within an assignable time interval. In the Radiological Health Handbook-1970 and ANSI N1.1-1976), coincidence counting is defined as ...”the occurrence of counts in two or more detectors simultaneously or within an assigned time interval. A true coincidence is one that is due to the incidence of a single particle or of several genetically related particles. An accidental, chance or random coincidence is one that is due to the accidental occurrence of unrelated counts in the separate detectors. An anticoincidence is the occurrence of a count in a specified detector unaccompanied simultaneously or in an assignable time interval by a count in other specified detectors. A delayed coincidence is the occurrence of a count in one detector at a short, but measurable, time after a count in another detector. The two counts are due to a genetically related occurrence such as successive events in the same nucleus.”

compliance - conformity or in accordance, as in compliance with rules, orders or guidance

confidence interval - A range of values for which there is a specified probability (e.g., 80%, 90%, 95%) that this set contains the true value of an estimated parameter.

data life cycle (DLC) - The process of planning, implementing, and assessing the survey plan and assessing the survey results prior to making a decision.

data qualifiers - qualitative designators (such as the letters U, J, R) used in the data verification and validation process that reflect the quality of the analytical data obtained during a remediation project. The data qualifiers reflect, in a qualitative manner, how the data meet the project objectives defined by the quality indicators. The quality indicators, when verifying and validating analytical data, are defined as: Measurable attributes of the attainment of the necessary quality for a particular environmental decision. Indicators of quality include precision, bias, completeness, representativeness, reproducibility, comparability and statistical confidence.

data quality assessment (DQA) - The process of assessing the survey results, determining that the quality of the data satisfies the objectives of the survey, and interpreting the survey results as they apply to the decision being made. This process focuses on the scientific and statistical evaluation of data to determine if the data are of the right type, quality, and quantity to support their intended use.

data quality objective (DQO) - A process to ensure that the survey results are of sufficient quality and quantity to support the final decision. It includes both qualitative and quantitative statements that clarify study objectives, define the appropriate type of data, and specifies levels of potential decision errors that will be used as the basis for establishing the quality and quantity of data needed to support decisions.

data validation - To substantiate, confirm, or approve the observed rule. Confirmation by examination and provision of objective evidence that the particular requirements for a specific intended use are fulfilled. In design and development, validation concerns the process of examining a product or result to determine conformance to user needs.

data verification - To ascertain the true value, authenticate, or establish as the correct value. Confirmation by examination and provision of objective evidence that specified requirements have been fulfilled. In design and development, verification concerns the process of examining a result of a given activity to determine conformance to the stated requirements for that activity.

$DCGL_W$ - The wide-area DCGL, which is based on pathway modeling. It is the uniform residual radioactivity concentration level within a survey unit that corresponds to the release criterion (e.g., regulatory limit in terms of dose or risk). This is the area used for the averaging of contamination to assess exposures to human and/or ecological receptors.

decommissioning - The process of removing a facility or site from operation; to retire (as in a facility) from active service. Specifically, in reference to MARSSIM, it is the process of removing a site safely from service, reducing residual radioactivity through remediation (decontamination) to a level that permits release of the property and termination of the license or other authorization for site operation. The objective of decommissioning is to reduce the residual radioactivity in structures, materials, soils, groundwater, and other media at the site so that the concentration of each radionuclide that contributes to residual radioactivity is indistinguishable from the background radiation concentration for that radionuclide.

derived concentration guideline level (DCGL) - Based on pathway modeling, this is the uniform residual radioactivity concentration level within a survey unit that corresponds to the release criterion (e.g., regulatory limit in terms of dose or risk). Again, based on pathway modeling, it is a radionuclide-specific predicted concentration or surface area concentration of specific nuclides that could result in a dose (TEDE - Total Effective Dose-Equivalent, or CEDE - Committed Effective Dose Equivalent) equal to the release criterion. In the MARSSIM, such concentration is termed the DCGL.

direct measurement - A type of measurement of radioactivity or radiation exposure which gives an immediate and direct response from which the quantity of radioactivity or radiation exposure can be inferred eiither through calibration of the instrument against a standard or by calculation.

dose - the quantity of radiation absorbed by a given mass of material or tissue. The unit of dose is the rad or gray (Gy).

double blind samples - A sampling and analysis procedure in which neither the sample-labeler nor the analytical chemist knows the source of the sample.

duplicate - One of two independent samples collected in such a manner that they are equally representative of the parameter(s) of interest at a given point in space and time.

dwell time - The duration over which a survey instrument is kept in one place before recording a radiation level. The shorter the dwell time, the less sensitive the measurement (i.e., higher detection limit).

elevated measurement - A measurement that exceeds a specified value $DCGL_{EMC}$.

elevated measurement comparison (EMC) - This comparison is used in conjunction with the Wilcoxon test to determine if there are any measurements that exceed a specified value, $DCGL_{EMC}$.

exposure - The state of being exposed, such as exposure to radiation, or to the elements. This is usually determined as a measure of x- or gamma-radiation at a certain place, based on its ability to produce ionization in air. The unit of exposure is the roentgen (R).

final status survey - Measurements and sampling to describe the radiological conditions of a site, following completion of decontamination activities (if any) and in preparation for release.

historical site assessment (HSA) - A detailed investigation to collect existing information, primarily historical information, on a site and its surroundings.

hot spot - A strong or intense level of radioactivity, above an investigation level, in a particular location on a site.

human factors efficiency - Of or pertaining to study of human proficiency in performance and the ability to accomplish a job with a minimum expenditure of effort; the ratio of work done or energy expended to accomplish a given task.
.

hypothesis - An assumption about a property or characteristic of a set of data under study. Tentative conclusion about the distribution function of a random variable. The goal of a statistical inference is to decide which of two complementary hypotheses is likely to be true. The null hypothesis describes what is assumed to be the true state of nature and the alternative hypothesis describes the opposite situation.

impacted area - Any area that is not classified as non-impacted. Areas with a possibility of containing residual radioactivity in excess of natural background or fallout areas.

instrument time constant - A measure of the response time of an instrument; it is the product of the resistance and capacitance of the electrical circuit.

investigation level - A derived media-specific, radionuclide-specific concentration or activity level of radioactivity that is based on the release criterion that triggers some response, such as a further investigation or cleanup, if the level is exceeded. May be used early in decommissioning to identify areas requiring further investigation, and may also be used as a screening tool during compliance demonstration to identify potential problem areas. A DCGL is an example of a specific investigation level. See also action level.

Less-than data - Measurements that are less than the minimum detectable concentration.

lower bound of gray region (LBGR) - The minimum value of the gray region. The width of the gray region (DCGL-LBGR) is also referred to as the shift, Δ.

matrix spike - A known amount of a specific radionuclide [e.g., an aliquot of sample with a known concentration of target analyte(s)] added to a sample of the matrix being analyzed (soil, water, filter, etc.), prior to sample preparation and analysis, to determine the fraction of the radionuclide present in the sample that is recovered by chemical separations or other sample handling processes. A matrix spike is used to document the bias of a method in a given sample composition. It may be appropriate to direct the laboratory to use specific samples as matrix spikes, and it may be necessary to provide additional samples for this purpose.

mean - a quantity having a value intermediate between the values of other quantities; an average which is the sum of the values of the observations divided by the number of observations.

median - the middle number in a given sequence, or the average of the two middle numbers when the sequence has an even number of numbers (e.g., 4 in the median of 1, 3, 4, 8, 9).

minimum detectable concentration (MDC) - The a-priori activity level that a specific instrument and technique can be expected to detect 95% of the time. When stating the detection capability of an instrument, this value should be used. The MDC is the detection limit, L_d, multiplied by an appropriate conversion factor to give units of activity.

non-parametric test - A test based on relatively few assumptions about the exact form of the underlying probability distributions of the measurements. As a consequence, nonparametric tests are generally valid for a fairly broad class of distributions. The Wilcoxon Rank Sum test and the Sign test are examples of nonparametric tests.

null hypothesis - An assumption that the distribution, value, or set of observations or measurements are not of the same distribution or source. The hypothesis that the distribution function is the expected one, and the discrepancy between expected and observed is due to chance.

outlier - A value which lies outside the range of other values in a set of observations; measurements that are unusually large relative to the bulk of the measurements in the data set.

performance evaluation - To carry through or execute in the proper or established manner in order to fulfill a command, promise or undertaking, such as the decommissioning of a site, and to determine or set the value or amount or to appraise and/or ascertain the numerical value of a function or relation, such as the effectiveness of a decommissioning action..

Poisson observation - An observation from a Poisson distribution. The Poisson distribution often describes the distribution of the number of events or objects on a fixed interval of time (e.g., counts per minute) or space (e.g., defects per square meter). It has a single parameter which corresponds to both the mean and variance of the distribution of counts.

power curve - The relationship between the number of samples to be taken and the ability to detect a significant difference between the sampled population and the comparison population.

precision - The extent to which a set of repeated measurements agree with one another; the degree of mutual agreement characteristic of independent measurements as the result of repeated application of the process under specific conditions. A set of perfectly precise measurements of the same quantity will all have the same value. It is concerned with the closeness of results. Note that precise measurements are not necessarily accurate; see the definition for bias.

quality assurance (QA) - An integrated system of management activities involving planning, implementation, assessment, reporting, and quality improvement to ensure that a process, item, or service is of the type and quality needed and expected by the client.

quality assurance/quality control (QA/QC) - Procedures and levels of documentation which are performed during implementation of a survey plan to collect information necessary to evaluate the survey results.

quality assurance project plan (QAPP) - The process which documents how quality assurance and quality control (QA/QC) are applied to obtain results that are of the type and quality needed and expected. This is usually manifested into a formal document describing in comprehensive detail the necessary QA, QC, and other technical activities that must be implemented to ensure that the results of the work performed satisfies the stated performance criteria.

quality control (QC) - The overall system of technical activities that measures the attributes and performance of a process, item, or service against defined standards to verify that they meet the stated requirements established by the client. The system includes operational techniques and activities that are oriented toward and used to fulfill requirements for quality.

quantile - (In a frequency distribution) - One of the values of a variable that divides the distribution of a variable into groups having equal frequencies (e.g., as a quartile which would divide the variable into four groups having equal frequencies).

receptor - an organ or group of organs exposed to stimulating agents (e.g., radionuclides).

reference (background) unit - A level which is considered as normal, ambient or natural, beyond which any addition is considered added by other means (e.g., anthropogenic - man-made means).

relative shift (Δ/α) - Δ divided by α, the standard deviation of the measurements.

release criterion - a regulatory limit expressed in terms of dose (mSv/y or mrem/y) or risk (cancer incidence or cancer mortality). The terms, release limit or cleanup standard have also been used to describe this term.

remediation - . The process and associated activities resulting in removal of contamination from a site. Remediation is sometimes used interchangeably with the terms remedial action, response action, or decontamination..

risk - The hazard or chance (degree of probability) of loss, exposure or injury.

roentgen (R) - The unit of radiation equal to the amount of x- or gamma-radiation that will produce ions in air containing a quantity of positive or negative electricity equal to one electrostatic unit in 0.001293 gram of air. See also exposure.

sample distribution function - Frequency of each value obtained in a random sample.

scanning - An evaluation technique performed by moving a detection device over a surface at a specified speed and distance above the surface to detect radiation.

scanning survey - The process of identifying contaminants (e.g., radionuclides) of concern and their relative abundances on a site or tract of land, building, structure, by scanning the surface with a detection device at a specified speed and distance above the surface to detect radiation. Note that in most multi-radionuclide contaminations, only sampling and lab analyses can (and not always "will") identify the radionuclides of concern and their relative abundances.

scoping survey - A type of survey that is conducted to identify (1) radionuclide contaminants, (2) relative radionuclide ratios, and (3) general levels and extent of contamination.

Sievert (Sv) - The special name for the International System (SI) unit of dose equivalent [1 Sv = 100 rem = 1 Joule per kilogram].

sign test - A nonparametric statistical test used to determine compliance with the release criterion when the radionuclide of interest is not present in background and the distribution of data is not symmetric. See also Wilcoxon Rank Sum test.

single blind sample - A sampling and analysis procedure in which the analytical chemist does not know the source of the sample.

site - Any installation, facility, or discrete, physically separate parcel of land, or any building or structure or portion thereof, that is being considered for survey and investigation.

skewed distribution - (Statistics) - A statistical distribution with a majority of the observations above or below the arithmetic mean. Such measurements or observations of values have an oblique course or distorted form; asymmetric, oblique direction or position about an axis.

surface soil - The outer face or exterior boundary of the soil, generally confined to the top six inches (15 centimeters).

survey unit - A geographic area consisting of structures and/or land areas of specified size and shape at a remediated site for which a separate decision will be made whether the unit attains the site-specific reference-based cleanup standard for the designated pollution parameter. Survey units are generally formed by grouping contiguous site areas with a similar use history and the same classification of contamination potential. Survey units are established to facilitate the survey process and the statistical analysis of survey data.

symmetric distribution - A statistical distribution with approximately equal numbers of observations above and below the arithmetic mean. Such sets of observations typically consist of pairs of points exhibiting symmetry about an axis, or a structure that exhibits a regular repeated pattern or symmetry of the component parts and a structure which is divisible into two similar parts by more than one plane passing through the center.

Type I decision error - The probability of rejecting the null hypothesis when it is true, or accepting the alternative hypothesis when it is false.

Type II decision error - The probability of accepting the null hypothesis when it is false.

uniform sample space - One for which all outcomes are equally likely.

vicinity property - An area or region adjacent to or near a place being examined, evaluated, remediated, decommissioned or otherwise surveyed. Vicinity properties may have been contaminated by any manner of transport mechanisms, including wind, water erosion, and groundwater contamination. For instance, in the case of inactive mill sites, it represents locations away from the mill sites where uranium mill tailings were used for construction or were transported off-site by wind or water erosion.

Wilcoxon Rank Sum (WRS) test - A nonparametric statistical test used to determine compliance with the release criterion when the radionuclide of concern is present in background. See also Sign test.

DISTRIBUTION LIST

Deputy Administrator
Assistant Administrators
EPA Regional Administrators
EPA Laboratory Directors
Deputy Assistant Administrator for Office of Policy, Planning and Evaluation
  Director, Office of Strategic Planning and Environmental Data
  Director, Office of Policy Analysis
  Director, Office of Regulatory Management and Evaluation
Deputy Assistant Administrator for Air and Radiation
  Director, Office of Radiation and Indoor Air
  Director, Office of Radiation Programs
EPA Headquarters Libraries
EPA Regional Libraries
EPA Laboratory Libraries
Library of Congress
National Technical Information Service
Congressional Research Service