March 30, 2022

EPA-SAB-22-001

The Honorable Michael S. Regan
Administrator
U.S. Environmental Protection Agency
1200 Pennsylvania Avenue, N.W.
Washington, D.C. 20460

Subject: Transmittal of the Science Advisory Board Report titled "Review of Multi-Agency Radiation Survey and Site Investigation Manual, Revision 2" (Public Comment Draft), dated May, 2020

Dear Administrator Regan,

Please find enclosed the final report from the Science Advisory Board (SAB). The EPA's Office of Air and Radiation requested that the SAB review the Multi-Agency Radiation Survey and Site Investigation Manual (MARSSIM), Revision 2 (Public Comment Draft). In response to the EPA's request, the SAB augmented the SAB Radiation Advisory Committee with subject matter experts to conduct the review.

The SAB Radiation Advisory Committee augmented for the review of the Multi-Agency Radiation Survey and Site Investigation (the MARSSIM Panel) held one teleconference on December 3, 2020, to deliberate on the agency's charge questions and met virtually on January 11-14, 2021, to deliberate on the responses to the agency's charge questions. The panel also met virtually on December 6, 2021, to deliberate its draft report. Oral and written public comments were considered throughout the advisory process. This report conveys the consensus advice of the panel.

While the SAB includes many recommendations within this report, we would like to highlight the following:

- Technological advancements for "scan-only" surveys are not adequately addressed in the draft revision to MARSSIM. To be technically appropriate and useful for performing environmental radiological surveys, statistical and uncertainty methodologies should be updated for modern detection systems with data logging. The SAB strongly encourages additional guidance development for scan-only methodologies.

- Regarding the use of statistics in MARSSIM, the SAB finds the study design concepts, methodologies and examples comprising MARSSIM-Revision 2 are indeed technically appropriate and highly useful. Much of the advice provided by the SAB involves the issue of clarity. The SAB recommends the inclusion of additional introductory material in which the concepts and terminology used throughout the MARSSIM manual are introduced early in the document.

- The SAB supports the inclusion of Scenario B (assumes the level of radioactive material in the survey unit meets the release criteria until proven otherwise), and its proposed implementation. The SAB agrees that it is reasonable to only recommend Scenario B when Scenario A (assumes the survey unit does not meet the release criteria) is infeasible (i.e., when the release criteria is close to zero); and a retrospective power analysis must be performed to prove the survey has sufficient statistical power to detect a survey unit that should not have passed.

- Many working examples are requested in order to assist in making the manual easier to understand.

- The SAB finds some shortcomings in the proposed implementation of the concept of measurement quality objectives (MQOs). Regarding uncertainty calculations, the SAB distinguishes three concepts, method uncertainty, required method uncertainty and measurement uncertainty. The first two of these are *a priori* concepts resting on predicted rather than observed data for a particular sample and measurement method; while the third is calculated *a posteriori* from the data observed during site investigation. All three of these require further clarification.

- The SAB finds the description of the concept of detection capability and its implementation in the draft MARSSIM document to be generally adequate and correctly described. The SAB recommends the Minimum Detectable Concentration (MDC) be evaluated with all known sources of uncertainty being properly quantified and combined into an appropriate expanded uncertainty. Equation 6-5 for calculation of MDC should be rewritten to include the uncertainty in C, the factor that converts the detection limit from blank/background counting signal to concentration measurement.

- The SAB agrees the Unity Rule should be maintained in MARSSIM Revision 2 for evaluation of multiple areas of elevated activity. The Unity Rule is used to ensure that the total dose (risk) from all sources and all radionuclides associated with each source does not exceed the release criteria. As an alternative, the dose or risk due to the actual residual radioactive material distribution could be calculated if an appropriate exposure pathway model is available.

- The SAB finds that MARSSIM content pertaining to survey sites containing discrete radioactive particles (DRPs) is inadequate and fails to address many important considerations. MARSSIM limits its discussion of DRPs to an appropriate and useful cautionary statement advising against using the Elevated Measurement Comparison (EMC) process when DRPs are discovered. However, the SAB believes the proposed rule of thumb to avoid using the EMC process may not prove useful or practical. The possible health risks posed by DRPs should be noted as they can be distinctly different from those caused by radioactive substances widely dispersed on building surfaces and within surface soil. In addition, MARSSIM should review how DRPs may present measurement challenges that will affect instrument selection and use, including the concerns regarding the mobility of DRPs and associated contamination hazards.

- The SAB finds the description of measurement methods and instrumentation information in Chapter 6 are generally useful and in large part appropriate and clear. However, recommendations for improvements in Chapter 6 are made to acknowledge when differences between ideal and realistic conditions merit specific treatment in the technical approaches; and to provide updated statistical techniques for modern data-logging systems that no longer rely on human surveyor data interpretation to perform the survey.

- The revised description in MARSSIM of how to set the Lower Bound of the Gray Region (LBGR) conveyed the point that LBGR should be set using site-specific information about the remaining residual contamination rather than some rule of thumb of a more general nature. The SAB agrees that setting the LBGR to a value near to the median seen in preliminary data is a good suggestion as long as the preliminary data are reasonably informative. In cases where preliminary data are limited, adherence to a heuristic rule probably can't be avoided.

As the EPA finalizes its draft Multi-Agency Radiation Survey and Site Investigation Manual (MARSSIM), Revision 2, the SAB encourages the Agency to address the panel's concerns raised in the enclosed report and consider their advice and recommendations. The SAB appreciates this opportunity to review the draft MARSSIM Revision 2 and looks forward to the EPA's response to these recommendations.

Sincerely,

/s/

Alison C. Cullen, Sc.D.
Chair
EPA Science Advisory Board

/s/

Daniel O. Stram, Ph.D.
Chair
EPA MARSSIM Review Panel

Enclosure:

NOTICE

This report has been written as part of the activities of the EPA Science Advisory Board, a public advisory committee providing extramural scientific information and advice to the Administrator and other officials of the Environmental Protection Agency. The Board is structured to provide balanced, expert assessment of scientific matters related to problems facing the Agency. This report has not been reviewed for approval by the Agency and, hence, the contents of this report do not represent the views and policies of the Environmental Protection Agency, nor of other agencies in the Executive Branch of the Federal government, nor does mention of trade names or commercial products constitute a recommendation for use. Reports of the EPA Science Advisory Board are posted on the EPA website at https://sab.epa.gov.

U.S. Environmental Protection Agency
Science Advisory Board
Radiation Advisory Committee Augmented for MARSSIM (Rev 2) Review

CHAIR
Dr. Daniel O. Stram, Professor, Department of Population and Public Health Sciences, Keck School of Medicine, University of Southern California, Los Angeles, CA

MEMBERS
Dr. Sally A. Amundson, Associate Professor of Radiation Oncology, Center for Radiological Research, Columbia University, New York, NY

Dr. Roland Benke, Director, Renaissance Code Development, Austin, TX

Dr. Harry M. Cullings, Consultant, Statistics, Radiation Effects Research Foundation (RERF), Pittsburgh, PA

Dr. Lawrence Dauer, Attending Physicist and Corporate Radiation Safety Officer, Memorial Sloan-Kettering Cancer Center, New York, NY

Mr. Earl W. Fordham, Deputy Director, Office of Radiation Protection, Division of Environmental Public Health, Washington Department of Health, Richland, WA

Dr. Eric Goldin, Project Manager, Goldin & Associates, Fallbrook, CA, United States

Ms. Barbara L. Hamrick, Radiation Safety Officer, University of California, Irvine Medical Center, Orange, CA

Dr. Kenneth G.W. Inn, Independent Consultant, Ewa Beach, HI

Dr. Robert Litman, Independent Consultant, The Villages, FL

Mr. Dennis Quinn, Owner, DAQ, Inc., Bethel, CT

Dr. Richard Smith, Professor, Department of Statistics and Operations Research, University of North Carolina, Chapel Hill, NC

Mr. Zoltan Szabo, Research Hydrologist, New Jersey Water Science, U.S. Geological Survey, Lawrence Township, NJ

Dr. Wei-Hsung Wang, Professor and Director, Radiation Safety Office, Louisiana State University, Baton Rouge, LA

Dr. R. Craig Yoder, Independent Consultant, Matthews, NC

CONTRIBUTORS.

Dr. Timothy DeVol, Professor, Environmental Engineering and Earth Sciences, Clemson University, Clemson, SC

Dr. Annie B. Kersting, Director of University Relations and Science Education, Lawrence Livermore National Laboratory, Berkeley, CA

Dr. Amy Kronenberg, Principal Staff Scientist, Life Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA

Dr. Brant Ulsh, Principal Health Physicist, M.H. Chew & Associates, Cincinnati, OH

SCIENCE ADVISORY BOARD STAFF

Dr. Diana Wong, Designated Federal Officer, U.S. Environmental Protection Agency, Washington, DC

U.S. Environmental Protection Agency
Science Advisory Board

CHAIR
Dr. Alison C. Cullen, Daniel J. Evans Endowed Professor of Environmental Policy, Evans School of Public Policy & Governance, University of Washington, Seattle, WA

MEMBERS
Dr. C. Marjorie Aelion, Associate Vice Chancellor for Research and Engagement and Professor of Environmental Health Sciences, University of Massachusetts Amherst, Amherst, MA

Dr. David T. Allen, Gertz Regents Professor of Chemical Engineering and Director of the Center for Energy and Environmental Resources, Department of Chemical Engineering, The University of Texas, Austin, TX

Dr. Susan Anenberg, Associate Professor, Department of Environmental and Occupational Health, Milken Institute School of Public Health, George Washington University, Washington, DC

Dr. Florence Anoruo, Assistant Professor of Plant and Environmental Science and Associate Research Scientist, Department of Biological and Physical Sciences, South Carolina State University, Orangeburg, SC

Dr. Joseph Arvai, Director of the Wrigley Institute for Environmental Studies and Dana and David Dornsife Professor of Psychology, Department of Psychology, University of Southern California, Los Angeles, CA

Dr. Barbara D. Beck, Principal, Gradient, Boston, MA

Dr. Roland Benke, Director, Renaissance Code Development, LLC, Austin, TX

Dr. Tami Bond, Scott Presidential Chair in Energy, Environment and Health, Department of Mechanical Engineering, Colorado State University, Fort Collins, CO

Dr. Mark Borsuk, Professor of Civil and Environmental Engineering, Pratt School of Engineering, Duke University, Durham, NC

Dr. Sylvie M. Brouder, Professor and Wickersham Chair of Excellence in Agricultural Research, Department of Agronomy, Purdue University, West Lafayette, IN

Dr. Jayajit Chakraborty, Professor, Department of Sociology and Anthropology, University of Texas at El Paso, El Paso, TX

Dr. Aimin Chen, Professor of Epidemiology, Department of Biostatistics, Epidemiology and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA

Dr. Amy Childress, Professor and Director of Environmental Engineering, Sonny Astani Department of Civil & Environmental Engineering, University of Southern California, Los Angeles, CA

Dr. Weihsueh Chiu, Professor, Department of Veterinary Integrative Biosciences, College of Veterinary Medicine and Biomedical Sciences, Texas A&M University, College Station, TX

Dr. Ryan Emanuel, Associate Professor, Nicholas School of the Environment, Duke University, Durham, NC

Mr. Earl W. Fordham, Deputy Director, Office of Radiation Protection, Division of Environmental Public Health, Washington Department of Health, Richland, WA

Dr. John Guckenheimer, Professor and Bullis Chair of Mathematics, Emeritus, Department of Mathematics, Center for Applied Mathematics, Cornell University, Ithaca, NY

Dr. Steven P. Hamburg, Chief Scientist, Environmental Defense Fund, Providence, RI

Dr. Marccus Hendricks, Assistant Professor of Urban Studies and Planning and Director of the Stormwater Infrastructure Resilience and Justice (SIRJ) Laboratory, Urban Studies and Planning Program, School of Architecture, Planning and Preservation and School of Engineering, University of Maryland-College Park, College Park, MD

Dr. Selene Hernandez-Ruiz, Chemistry Program Manager, Colorado Department of Health and Environment, Lakewood, CO

Dr. Elena G. Irwin, Distinguished Professor of Food, Agricultural and Environmental Sciences in Economics and Sustainability and Faculty Director for the Sustainability Institute, Department of Agricultural, Environmental, and Development Economics, The Ohio State University, Columbus, OH

Dr. David Keiser, Professor, Department of Resource Economics, University of Massachusetts Amherst, Amherst, MA

Dr. Mark W. LeChevallier, Principal, Dr. Water Consulting, LLC, Morrison, CO

Dr. Angela M. Leung, Clinical Associate Professor of Medicine, Department of Medicine, Division of Endocrinology, Diabetes, and Metabolism, David Geffen School of Medicine; VA Greater Los Angeles Healthcare System, University of California Los Angeles, Los Angeles, CA

Ms. Lisa Lone Fight, Director, Science, Technology, and Research Department, MHA Nation, Three Affiliated Tribes, New Town, ND

Dr. Lala Ma, Assistant Professor, Department of Economics, Gatton College of Business and Economics, University of Kentucky, Lexington, KY

Dr. John Morris, Board of Trustees Distinguished Professor Emeritus, University of Connecticut, Ellington, CT

Dr. Enid Neptune, Associate Professor of Medicine, Department of Medicine, Division of Pulmonary and Critical Care Medicine, Johns Hopkins University, Baltimore, MD

Dr. Sheila Olmstead, Professor of Public Affairs, Lyndon B. Johnson School of Public Affairs, The University of Texas at Austin, Austin, TX

Dr. Austin Omer, Associate Director of Natural Resource Policy, Governmental Affairs and Commodities Division, Illinois Farm Bureau, Bloomington, IL

Dr. Gloria Post, Research Scientist, Division of Science and Research, New Jersey Department of Environmental Protection, Trenton, NJ

Dr. Kristi Pullen-Fedinick, Chief Science Officer, Natural Resources Defense Council, Washington, DC

Dr. Amanda D. Rodewald, Garvin Professor and Senior Director of Center for Avian Population Studies, Department of Natural Resources and the Environment, Cornell Lab of Ornithology, Cornell University, Ithaca, NY

Dr. Emma J. Rosi, Senior Scientist, Cary Institute of Ecosystem Studies, Millbrook, NY

Dr. Jonathan M. Samet, Dean and Professor, Departments of Epidemiology and Environmental and Occupational Health, Office of the Dean, Colorado School of Public Health, Aurora, CO

Dr. Elizabeth A. (Lianne) Sheppard, Rohm and Haas Professor in Public Health Sciences, Department of Environmental & Occupational Health Sciences and Department of Biostatistics, Hans Rosling Center for Population Health, University of Washington, Seattle, WA

Dr. Drew Shindell, Nicholas Distinguished Professor of Earth Science, Duke Global Health Initiative, Nicholas School of the Environment, Duke University, Durham, NC

Dr. Genee Smith, Assistant Professor, Department of Environmental Health and Engineering, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD

Dr. Richard Smith, Professor, Department of Statistics and Operations Research, University of North Carolina, Chapel Hill, NC

Dr. Daniel O. Stram, Professor, Department of Population and Public Health Sciences, Keck School of Medicine, University of Southern California, Los Angeles, CA

Dr. Peter S. Thorne, Professor and Head, Department of Occupational & Environmental Health, and Director of Human Toxicology Program, Department of Occupational & Environmental Health, College of Public Health, University of Iowa, Iowa City, IA

Dr. Godfrey Arinze Uzochukwu, Senior Professor, Waste Management Institute, North Carolina Agricultural and Technical State University, Greensboro, NC

Dr. Wei-Hsung Wang, Professor, Center for Energy Studies and Director of the Radiation Safety Office, Louisiana State University, Baton Rouge, LA

Dr. June Weintraub, Senior Epidemiologist and Manager of Water and Noise Regulatory Programs, San Francisco Department of Public Health, San Francisco, CA

Dr. Sacoby Wilson, Associate Professor and Director of the Center for Community Engagement, Environmental Justice, and Health (CEEJH), Maryland Institute for Applied Environmental Health, School of Public Health, University of Maryland-College Park, College Park, MD

Dr. Dominique van der Mensbrugghe, Research Professor and Director of the Center for Global Trade Analysis, Department of Agricultural Economics, Purdue University, West Lafayette, IN

SCIENCE ADVISORY BOARD STAFF

Dr. Thomas Armitage, Designated Federal Officer, U.S. Environmental Protection Agency, Washington, DC

TABLE OF CONTENTS

ABBREVIATIONS AND ACRONYMS

AL	Action Level
ANOVA	Analysis of Variance
ANSI	American National Standards Institute
CSU	Combined standard uncertainty
CV	Coefficient of Variation
DCGL	Derived Concentration Guideline Level
$DCGL_{EMC}$	Derived Concentration Guideline Level for elevated measurement comparison
$DCGL_{w}$	Derived Concentration Guideline Level for wide area
DIS	Direct Ion Storage
DL	discrimination limit
DOD	U.S. Department of Defense
DOE	U.S. Department of Energy
DQO	Data Quality Objectives
DRPs	discrete radioactive particles
ED	Electronic Dosimeters
EIC	electret ion chamber
EMC	Elevated Measurement Comparison
FSS	Final status survey
GPS	global positioning system
HSA	Historical Site Assessment
HTD	hard-to-detect
IEEE	Institute of Electrical and Electronics Engineers
ISO	International Organization for Standardization
L_C	Critical Level
L_D	Detection Limit
LBGR	Lower Bound Gray Region
LSC	Liquid Scintillation Counter
MARLAP	Multi-Agency Radiological Laboratory Analytical Protocols Manual
MARSAME	Multi-Agency Radiation Survey and Assessment of Materials and Equipment
MARSSIM	Multi-Agency Radiation Survey and Site Investigation Manual
MDA	Minimum Detectable Activity
MDC	Minimum Detectable Concentration
MDCR	Minimum Detectable Count Rate
MDER	Minimum Detectable Exposure Rate
MQC	Minimum Quantifiable Concentration
MQO	Measurement Quality Objectives
NIST	National Institute of Standards and Technology
NJDEP	New Jersey Department of Environmental Protection
NRC	Nuclear Regulatory Commission
ORAU	Oak Ridge Associated Universities
OSL	optically stimulated luminescence
PIC	pressurized ionization chamber
PNNL	Pacific Northwest National Laboratory
RSS	Ranked Set Sampling
SAB	Science Advisory Board
SOPs	Standard Operating Procedures
SRS	Simple Random Sampling

TLD	thermoluminescent dosimeter
UBGR	Upper Bound Gray Region
UCL	upper confidence limit
U.S. EPA	U.S. Environmental Protection Agency
WMW	Wilcoxon-Mann-Whitney
WRS	Wilcoxon Rank Sum
WSR	Wilcoxon signed rank
WT	Wilcoxon Test

SYMBOLS, NOMENCLATURE, and NOTATIONS

$<$	less than
$>$	greater than
$\leq$	less than or equal to
$\geq$	greater than or equal to
%	percent
$1-\beta$	statistical power of a hypothesis test
α	Type I decision-error rate
α_Q	alpha used for the quantile test
A	area
β	Type II decision-error rate
b	background count rate
B	mean background counts
Bq	becquerel
C	radionuclide concentration or activity
C	constant
Ci	curie
C_i	concentration value an individual radionuclide (i = 1, 2, …, n)
$c_i\mu(x_i)$	component of the uncertainty in y due to x_i
cm	centimeter
cm^2	square centimeter
cm^3	cubic centimeter
cpm	counts per minute
Δ	shift (width of the gray region, UBGR–LBGR)
Δ/σ	relative shift
dpm	disintegrations per minute
ft	foot (feet)
ft^3	cubic foot (feet)
g	gram
GBq	gigabecquerel (1×10^9 becquerels)
h	hour
H_0	null hypothesis
H_1	alternative hypothesis
i	ith sample or measurement in a set
in	inch
k	k-statistic for the quantile test
k	coverage factor for the expanded uncertainty, U
k	Poisson probability sum for α and β (assuming α and β are equal)
k	critical value of the sign test
kBq	kilobecquerel (1×10^3 becquerels)
kg	kilogram
km	kilometer
k_Q	multiple of the standard deviation defining y_Q, usually chosen to be 10
L	length
L	liter

L_C	critical level
L_D	detection limit
lb	pound
µ	micro (10^{-6})
µ	true mean
μ	theoretical mean of a population distribution
µBq	microbecquerel
µCi	microcuries
µR	microroentgen (1×10^{-6} roentgen)
µSv	microsievert
m	number of reference measurements (WRS test or Quantile test)
m	meter
m^2	square meter
M_i	total amount of [dose counts, activity, etc.]
mBq	millibecquerels
$MDCR_{surveyor}$	required number of net source counts
mg	milligram(s)
mGy	milligray
mm	millimeter(s)
mR	milliroentgen
mrad	millirad
mrem	millirem (1×10^{-3} rem)
mSv	milliseivert (1×10^{-3} Sv)
n	number of survey unit measurements (WRS test or Quantile test)
N	sample size (i.e., number of data points [or samples]) for the Sign test
p	coverage probability for expanded uncertainty
P	probability
pCi	picocurie (1×10^{-12} curies)
ppt	parts per trillion
$\rho(X_i,X_j)$	correlation coefficient for two input quantities, X_i and X_j
R	ratio
R	roentgen (exposure rate)
Ra	radium (isotopes listed: ^{224}Ra, ^{226}Ra, ^{228}Ra)
R_B	mean background count rate
R_{net}	net counting rate
$r(x_i,x_j)$	correlation coefficient for two input estimates, x_i and x_j
σ	theoretical total standard deviation of the population distribution being sampled
$\hat{\sigma}$	the standard deviation of the measured analyte
σ^2	theoretical total variance of the population distribution being sampled
σ_M	theoretical measurement standard deviation of the population distribution being sampled, estimated by the combined standard uncertainty of the measurement
σ_n	standard deviation of the net count rate result
σ_M^2	theoretical measurement variance of population distribution being sampled
σ_{MR}	required measurement method standard deviation (upper limit)

$\sigma(X_i, X_j)$	covariance for two input quantities, X_i and X_j
σ_s	estimate of the measurement variability in the survey unit
σ_y	total uncertainty
Sr	strontium (isotope listed: ^{90}Sr)
Sv	seivert
t	t-test statistic
$t_{1/2}$	half-life
Tc	technetium (isotopes listed: ^{99}Tc, ^{99m}Tc)
Th	thorium (isotopes listed: ^{228}Th: ^{230}Th, ^{232}Th, ^{234}Th)
U	expanded uncertainty
U	uranium (isotopes listed: ^{234}U, ^{235}U, ^{238}U)
$u(x_i)$	standard uncertainty of the input estimate, x_i
$u(x_i,x_j)$	covariance of two input estimates, x_i and x_j
$u_c(y)$	combined standard uncertainty of y
u_M	measurement method uncertainty
u_{MR}	required measurement method uncertainty
$\widehat{\omega}$	standard deviation of background variability
x	estimate of the input quantity, X
x_i	results of the individual samples
$z_{1-\alpha}$	$(1-\alpha)$-quantile of the standard normal distribution
y	year

1. INTRODUCTION

The Environmental Protection Agency (EPA) Office of Air and Radiation (OAR) requested that the Science Advisory Board (SAB) conduct a peer review of the technical accuracy and understandability of its document titled "Multi-Agency Radiation Survey and Site Investigation Manual (MARSSIM), Revision 2 (Draft for Public Comment)" (hereafter referred to as the draft MARSSIM). MARSSIM was developed by the technical staff of the four Federal agencies having authority for control of radioactive materials: the Department of Defense (DOD), the Department of Energy (DOE), EPA, and the Nuclear Regulatory Commission (NRC). It provides information on planning, conducting, evaluating, and documenting environmental radiological surveys of surface soil and building surfaces for demonstrating compliance with regulations.

In response to the EPA's request, the SAB convened a panel of subject matter experts to conduct the review. The Science Advisory Board Radiation Advisory Committee augmented MARSSIM Review Panel (hereafter referred to as the Panel) convened 3 public meetings to conduct a peer review of the EPA's revised document. Meetings were held on December 3, 2020, to discuss the charge questions, and January 11-14, 2021, to deliberate responses to charge questions. The Panel also met on December 6, 2021, to deliberate its draft report. Oral and written public comments were considered throughout the advisory process. Charge questions were specified by OAR on the changes made to MARSSIM Revision 1 (U.S. EPA, 2000) in the draft document and appendices. There are 3 main charge questions and 12 sub-questions.

This report is organized to state each charge question raised by the Agency followed by the consensus response and recommendations. The SAB provided key recommendations that are necessary to improve the critical scientific concepts, issues, and/or narrative within the EPA's document. The SAB also provided suggestions to strengthen the scientific concepts, issues and/or narrative within the document, but other factors (e.g., Agency need) should be considered by the Agency before undertaking these revisions.

A list of acronyms and abbreviations can be found at the front of this report to assist in orienting the reader to the terms and names used in the EPA's report and throughout the SAB's responses to the Charge Questions. EPA's charge questions to the panel can be found in Appendix A. All editorial comments are presented within Appendix B. Additional comments can be found in Appendix C. All materials and comments related to this report are available at:

https://sab.epa.gov/ords/sab/f?p=100:18:7282117609671:::RP,18:P18_ID:2582

2. RESPONSES TO EPA'S CHARGE QUESTIONS

2.1. Charge Question 1. Concepts and Methodologies

Are the revisions to MARSSIM concepts and methodologies technically appropriate, useful and clear, and do they provide a practical and implementable approach to performing environmental radiological surveys of surface soil and building surfaces?

Technological advancements for "scan-only" surveys are not adequately addressed in the draft MARSSIM. To be technically appropriate and useful for performing environmental radiological surveys, statistical and uncertainty methodologies should be updated for modern detection systems with data logging. The SAB recommends revision of the term "scan-only" surveys into generic scanning surveys and site-specific scanning surveys. Guidance should be prepared and presented for the revised definition. In particular, site-specific calibration for field surveys should be addressed and include explicit description of measurement quality objectives as well as commutability[1] of reference materials. The SAB recommends EPA justify the assignment of the minimum detectable concentration for scan-only surveys (Scan MDC) at 50% of the Derived Concentration Guideline Level for wide areas ($DCGL_W$). Nonstatistical uncertainties should be emphasized in the revised guidance.

Regarding the use of statistics in the draft MARSSIM, the issues that surround statistical study design are many and complex, and it is not possible to provide advice on the development of site surveys that is simultaneously useful and appropriate, technically sound and comprehensive, while also being jargon-free and easy to understand on a first read through by a non-statistically trained user. Overall, the SAB finds that the study design concepts, methodologies and examples comprising the draft MARSSIM are indeed technically appropriate and highly useful, and the points where the SAB disagrees with the advice given by the draft MARSSIM are mostly relatively small. Much of the advice provided by the SAB (especially for charge question 1.2) is on the issue of clarity. In particular, the SAB recommends the inclusion of additional introductory material in which the concepts and terminology used throughout the MARSSIM manual (and indeed throughout the Multi-Agency Radiological Laboratory Analytical Protocol (MARLAP, U.S. EPA, 2004), the Multi-Agency Radiation Survey and Assessment of Materials and Equipment Manual (MARSAME, U.S. EPA, 2009), and related documents and appendices be introduced early in the document. The SAB has provided some detail on what ground this introductory material should cover. The purpose of the proposed introduction is to flatten the learning curve faced by even statistically well-trained users, when initially confronted with MARSSIM terminology and specialized concepts.

[1] The general definition of commutability is "equivalence." Commutability is defined as property of a given reference material, demonstrated by the closeness of agreement between the relation among the measurement results for a stated quantity in this material, obtained according to two measurement procedures, and the relation obtained among the measurement results for other specified materials (ISO, 2012).
NOTES 1 The material in question is usually a calibrator. 2 At least one of the two given measurement procedures is usually a high-level measurement procedure.

Among the many points of agreement with the changes in the draft MARSSIM are the acceptance of Scenario B as a valid methodology so long as retrospective power analysis shows the design had adequate power to reject the null hypothesis as specified in the DQO. The SAB also agrees that it is reasonable to require that Scenario A (assumes the survey unit does not meet the release criteria) be preferred to Scenario B (assumes the level of radioactive material in the survey unit meets the release criteria until proven otherwise), except when Scenario A is infeasible (i.e., because the proposed number of samples for analysis becomes unreasonably large as the $DCGL_W$ approaches zero).

The SAB also agrees that the suggested sequence of tests for existence of background variability, first using the Kruskal-Wallis test, and then ANOVA, is reasonable. However, among the places where there is disagreement is in the use of a particular value, i.e., 3 $\hat{\omega}$ (3 standard deviation), to allow for background variability in the construction of tests under Scenario B, which the SAB finds to be overly forgiving. In addition, the SAB agrees with the use of the Quantile test as part of checking the assumptions of Scenario B but asks that additional information about the power of this test be provided in the manual.

The SAB found some shortcomings in the proposed implementation of the concept of measurement quality objective (MQO). Regarding uncertainty calculations, the SAB distinguishes three concepts, method uncertainty, required method uncertainty and measurement uncertainty. The first two of these are *a priori* concepts resting on optimal instrument or laboratory detection capabilities rather than observed data for a particular sample and measurement method; while the third is calculated *a posteriori* from the data observed during site investigation. All three of these require further clarification, for example, the calculation of required method uncertainty, which is an upper limit for the model uncertainty (at the Upper Bound Gray Region, UBGR) based on tolerable error rates and the width of the gray region, needs further detail and inclusion in an example. In calculation of overall measurement uncertainty, the description should provide examples of the known influences and whether the resultant uncertainty from these influences is Type A (derived from statistical methods) or Type B (usually non-statistical, like experience or expert knowledge) and the specific means by which their quantitative uncertainties (estimates of the standard deviation) were derived.

The SAB found the description of the concept of detection capability and its implementation in the draft MARSSIM document to be generally adequate and correctly described. A Minimum Detectable Concentration (MDC) should be evaluated with all known sources of uncertainty being properly quantified and combined into an appropriate expanded uncertainty. For example, treating calibration factors or surveyor efficiencies as known constants with no uncertainty is not appropriate and conflicts with National Institute of Science and Technology (NIST) and International Organization for Standardization (ISO) advice. At present the main comment about error in calibration (or other systematic errors that affect all measurements) is that such errors should be minimized through sound laboratory practice. This seems only partly adequate and formal incorporation of systematic errors into uncertainty propagation should be considered. The SAB also suggests that changes needed in Equation 6-18 (error propagation) to deal with correlated errors either of Type A or Type B should be described.

Regarding uncertainties in scanning and other field measurements, the SAB is in good agreement with what is described in Chapter 6 of the draft MARSSIM as far as they go, but there is an ambiguity about the advice regarding using these measurements for quantification as opposed to detection of radiation anomalies; it is explicitly stated in Chapter 6 that the scanning is only for detection (page 6-33) but this is partially contradicted by advice given on the next page. The SAB recommends that the MARSSIM-Revision 2 address more fully the topic of using scanning (or other field measurements) for quantification of average concentrations over a site. The other topic addressed in the SAB review of Chapter 6 is treatment of systematic uncertainties, especially since scanning a large fraction of a site may shrink random uncertainties down to negligible levels.

Regarding stakeholder complaints that calculating uncertainties for field measurements makes the survey process difficult to implement, the SAB notes that quantifying measurement uncertainty for field measurements (scan-only and *in situ*) is important, most especially for documentation of final status survey (FSS) results. A detailed list of factors affecting the performance of scan-only designs is provided in Chapter 6 and would seem to be useful to the stakeholders. It is noted on page 6-27 of the draft MARSSIM that rigorous uncertainty assessment for field measurements is generally only necessary for final site survey documentation, and generally not required in scoping or characterization surveys.

The SAB agrees with keeping the original MARSSIM requirement of a measurement method with MDC equal to between 10 and 50% of the UBGR so long as they relate to the data quality objectives, e.g., the tolerable error rates and the width of the gray region. The SAB prefers that the range (10 to 50% of the UBGR) be given, as in the original MARSSIM, rather than a single number (e.g., 50%) as in the draft MARSSIM. The SAB notes that this range is nearly equivalent to setting the minimum quantifiable concentration (MQC) to about 10 percent of the action level (at *alph*a = *beta* = 0.05). Therefore the "standard" choices for both MDC and MQC nearly concur, so that either could work well as an MQO. The SAB agrees with MARLAP (U.S. EPA, 2004) (Chapter 3 page 3-12 to 3-13) that the value of including the MQC as a possible performance characteristic is to emphasize the importance of the quantification capability of a method for those instances where the issue is not whether an analyte is present or not, but rather how precisely the analyte can be quantified.

The SAB finds the MARSSIM content pertaining to environmental radiological surveys of surface soil and building surfaces to be generally logical, technically appropriate, useful, but somewhat ambiguous. The draft MARSSIM provides a practical and implementable approach to performing environmental radiological surveys of surface soils and building surfaces. The SAB finds the elevated measurement comparison (EMC) approach for small areas of elevated concentrations of radioactive material acceptable. The SAB agrees with the decision to maintain the use of the Unity Rule for multiple areas of elevated activity. The second alternative cited in the draft MARSSIM would always be an option, providing that it can be implemented using sufficient characterization data about the exposure pathways of interest. Additional clarity is possible by adding more details in the body of the sections and the examples (e.g., degree of uncertainty, reasoning for the initial number of samples).

The SAB finds that the draft MARSSIM content pertaining to discrete radioactive particles (DRPs) is inadequate and fails to address many important considerations. MARSSIM limits its discussion of DRPs to an appropriate and useful cautionary statement advising against using the Elevated Measurement Comparison (EMC) process when DRPs are discovered; although the SAB believes the proposed rule of thumb to avoid using the EMC process may not prove useful or practical. The absence of critical information pertaining to the influence DRPs exert on the development of data quality objectives, measurement quality objectives and operational radiation safety programs impedes the ability of MARSSIM to provide practical and implementable approaches to performing environmental radiological surveys of surface soil and building surfaces. As part of MARSSIM's introductory sections, the possible health risks posed by DRPs should be noted as they can be distinctly different from those caused by radioactive substances widely dispersed on building surfaces and within surface soil. Some DRPs may cause injury from short-term, acute exposure if the particle comes into direct contact with tissues while other DRPs may pose little to no risk above that from more widely dispersed radioactive materials. MARSSIM should explicitly address through decision aids how such variability in risk may be considered during early site assessments and survey planning objectives. Additionally, MARSSIM should review how DRPs may present measurement challenges that will affect instrument selection and use, including the concerns regarding the mobility of DRPs and associated contamination hazards.

2.1.1. Charge Question 1.1. Scan Only Surveys

Please identify whether the inclusion and proposed implementation of scan-only surveys (Section 5.3.6.1 and Section 8.5) is appropriate, adequate, and clear, especially the discussion on sampling for scan-only measurement method validation or verification.

Consistent with the charge, the response is communicated according to four considerations:

- Technical appropriateness of concepts and methodologies
- Practicality of implementing approaches for surveys of surface soil and building surfaces
- Usefulness and clarity of MARSSIM revisions including misinterpretation of "scan-only"
- Verification of scan-only surveys with sampling and laboratory analysis

<u>Technical appropriateness of concepts and methodologies</u>

The SAB agrees that the current MARSSIM guidance does not adequately address modern scanning surveys. Arising from significant technological advancement over the past two decades, newer scanning instruments and mobile systems represent attractive options for consideration and assessment. Quantitative measurements with various example systems are described in the scientific literature (Marques et al. 2021; Peeva, 2021; Ji et al. 2020; Rahman et al. 2020; Ji et al. 2019; Lee and Kim, 2019; Sanada et al. 2019; Azami et al. 2018; Falciglia et al. 2018; Wilhelm et al. 2017; Sinclair et al. 2016; Sanada and Torii, 2015; Kock et al. 2014; Sanderson, 2013; Tanigaki et al. 2013; Kock and Samuelsson, 2011). Detection efficiency and minimum detectable activity for mobile scanning have been investigated regarding scanning speed and signal processing (Falkner and Marianno, 2021; Marianno, 2015). The SAB does not endorse specific detection systems or commercial equipment. Although commercial systems are not referenced in this report, the SAB reinforces the importance of detection system calibration that yields measurement quantification with uncertainties to support defensible final survey results.

The SAB strongly encourages additional guidance development for scan-only methodologies. The SAB finds additional discussion on instrument response and calibration factors for the specific site or surface being measured to be essential. Guidance is needed on site-specific calibration for field surveys, and the guidance should explicitly describe Measurement Quality Objectives (MQO) implementation and the commutability of reference materials in the context of scan-only surveys. Commutability has been defined by the International Organization for Standardization (ISO, 2015a) and the National Institute of Standards and Technology (NIST, 2020). In this context, commutability relates to the equivalence of measurement results for the reference material and field survey results of the same measured quantity by two or more measurement procedures. The SAB acknowledges that environmental factors and heterogeneities at the site, including variations in surface and instrument efficiency during scanning, can dominate measurement uncertainty of scan-only surveys. As a result, nonstatistical uncertainties [see Section 2.1.3.1, also referred to as Type B uncertainties by the National Institute of Standards and Technology (NIST, 1994)] should be emphasized. Relevant uncertainties should be clearly annotated while establishing data and measurement quality objectives, selecting survey methods and equipment, performing and confirming calibrations, conducting surveys, recording results, and generating preliminary and final conclusions. Modern data logging pairs detector position and raw instrument readout during continuous operation. Recording detector position during scanning measurements enables the generation of two-dimensional (or three-dimensional) maps of detector readout over the scanned surface. For scan-only surveys without data logging, the SAB finds the statistical methods in Section 5.3.6.1 and Section 8.5 of the draft revision to be technically appropriate with the technical improvements recommended in the next two paragraphs.

The SAB recommends that the EPA justify subjecting the minimum detectable concentration for scan-only surveys (Scan MDC) to 50% of the Derived Concentration Guideline Level for wide areas ($DCGL_W$). This justification should be in harmony with revised guidance on incorporating nonstatistical (Type B) uncertainties into MDC calculations for scan-only surveys, especially if conservative parameter values are intentionally selected to overestimate concentrations in lieu of quantitative uncertainty propagation. A more complete description of scan-only technology advances since the last revision of MARSSIM would facilitate guidance development. For example, addressing overlapping fields of view for scan-only surveys with gamma-ray detectors is advised.

The SAB supports development of a stronger, more formal connection between scan-only measurements and laboratory verification results to substantiate quantification requirements for scan-only surveys. Relationships should be explained among MDC recommendations for scan-only surveys, instrument calibration including its rigor for the specific site under investigation, and statistical approaches as well as the extent of conservatism in metrics selected for decision-making. In general, the SAB suggests EPA perform a consistency check of MDC recommendations relative to guideline levels throughout the MARSSIM documentation and explain the rationale for differences.

<u>Practicality of implementing approaches for surveys of surface soil and building surfaces</u>

The SAB recommends that the EPA present appropriate quality assurance and quality control definitions for the scanning technology and its recording output. Because the two-staged technique on scanning with frequent stationary measurements was designed for older equipment without data logging, measurement techniques for modern systems with continuous data logging of detector location and response should be described. The SAB also supports the incorporation of radiation data mapping generated by continuous data-logging scanning systems into guidance. On strengthening the description of relevant equipment for scan-only surveys, the SAB encourages EPA to describe how scan-only equipment relates to data and measurement quality objectives including measurement sensitivity requirements, to provide additional detail on how to perform scan-only surveys, to illustrate scan-only implementation with validation from sampling and laboratory analysis, and to present implementation examples of replicate measurements over defined scanning areas. Adding detailed insights from a case study is also advised.

Usefulness and clarity of MARSSIM revisions including misinterpretation of "scan-only"
The SAB recommends revising the term "scan-only" surveys into "generic scanning surveys" and "site-specific scanning surveys." Generic scanning surveys rely on systems with no site-specific calibration and thus require validation from sampling and laboratory analysis. Site-specific scanning surveys utilize systems that apply site-specific calibrations with commutable reference validation materials. When the implementer confirms the measurement quality objectives associated with calibration are met inclusive of nonstatistical (Type B) uncertainties, site-specific scanning systems require minimal to no additional confirmation from sampling and laboratory analysis. Slight adjustments to the organizational structure (e.g., Sections 5.3.6.1 and 5.3.6.2) would be expected to accommodate the revised definitions.

Expectations associated with the recommended percentages of areas covered by scan-only surveys should be clarified. Guidance on how to address 90% of the area scanned for Class 1 limited by obstacles in comparison to the 100% recommendation in current guidance could serve as one example. The SAB supports a discussion on surface versus volumetric contamination in the context of scan-only surveys. Per the bulleted consideration of alpha and beta radiation impacts on scan-only surveys in Section 5.3.6.1, the SAB agrees with expanding the description of scan-only survey impacts due to different particle emissions (e.g., alpha, beta, and gamma rays). Additionally, it would be helpful to incorporate examples with readouts or data logging results representative of modern scan-only technology. It is important to include nonstatistical (Type B) uncertainties in these examples. In particular, Example 6 in the draft MARSSIM Chapter 6 for scan MDC would include nonstatistical uncertainties in the surface efficiency and instrument efficiency. Revision 1 of NUREG-1507 (NRC, 2020) recommends "experimentally determined surface efficiencies for anticipated field conditions." In the draft MARSSIM, the surface efficiency is also referred to as the source efficiency. To account for natural variations in the field, surface efficiencies would be experimentally determined at more than one location. Uncertainty in the surface efficiency at the site would address these important nonstatistical effects. Uncertainty in the instrument efficiency would account for differences in source geometry and distribution. The SAB advises EPA to justify why final status surveys based on scan-only measurements that are consistent with the project MQOs would be inadequate. The SAB also supports the cautionary notes on applying statistical tests to small data sets and endorses extending these notes into other areas of the document by illustrating the risks from small data sets with examples.

Verification of scan-only surveys with sampling and laboratory analysis
The SAB maintains that collecting samples for laboratory analysis to validate scan-only measurements is a good quality control practice and recommends that the measurement quality

objective for laboratory analysis have a measurement uncertainty of at least one-third of that for the scan-only measurements [i.e., laboratory measurement uncertainties are at least 3 times smaller compared to scan-only measurement uncertainties (ANSI, 1995, see Appendix C)]. As clarified by the recommended revision to the term "scan-only," the SAB acknowledges site-specific instrument calibration, with commutable reference validation standards and measurement uncertainties consistent with the project MQOs for the contaminants under investigation, can minimize requirements for subsequent confirmation by physical sampling and laboratory analysis. Expanding the verification discussion for scan-only surveys to include spatial concentration variability, contamination depth, and anticipated contaminant migration is also advised.

Revised MARSSIM guidance should highlight important distinctions between scanning with stationary measurements versus scanning measurements with sampling for laboratory analysis. The SAB recommends a protocol for data-logging scanning systems with (1) site-specific instrument calibration, (2) preselected locations for stationary measurements according to a grid, (3) continuous scanning measurements at a constant speed, and (4) follow-up stationary measurements at locations suspected to have the highest concentrations. From a broader programmatic perspective, the SAB suggests EPA motivate and organize multiagency activities on scan-only performance testing to validate capabilities of modern systems and techniques similar to the performance specifications and testing methods described in N42.23 consensus standards by the American National Standards Institute and Institute of Electrical and Electronics Engineers (ANSI/IEEE, 2021).

Recommendations:

- Revise the term "scan-only" surveys into generic scanning surveys and site-specific scanning surveys. Develop guidance for the revised definition.
- Address site-specific calibration for field surveys and explicitly describe measurement quality objectives as well as commutability of reference materials in the context of scan-only surveys.
- Justify subjecting the minimum detectable concentration for scan-only surveys (Scan MDC) to 50% of the Derived Concentration Guideline Level for wide areas ($DCGL_W$).
- Emphasize uncertainties derived from other than statistical means (also referred to as Type B uncertainties) in the revised guidance for scan-only surveys.

2.1.2. Charge Question 1.2. Scenario B

General Comments

There are many parts of the draft MARSSIM Revision 2 manual (the draft MARSSIM, U.S. EPA, 2020) that are initially extremely challenging to first-time readers, even those with strong statistical training. This is partly because familiar concepts (to a statistician) are given new names, such as the whole pantheon of action level (AL), discrimination limit (DL), gray region, upper bound gray region (UBGR), and lower bound gray region (LBGR). There are some cases where this is compounded by problems of clarity, such as the use of key concepts or acronyms before they are defined. The gray region is an example, much discussion of the upper bound and lower bounds of the gray region under both Scenario A and B, appears throughout the document (indeed the "gray region" is cited 45 times total) but only in the glossary in the Appendix is a formal definition of the gray region provided. A discrepancy between the glossary and the text is in the definition of the lower bound of the gray region (LBGR) under Scenario B; the glossary defines this as being equal to the derived concentration guideline level for average concentrations over a wide area (DCGLw) whereas the text tends to call this the action level (AL). Once some familiarity is gained with the terminology, then the logic of the recommended procedures, statistical tests, null and alternative hypotheses, power, and sample size calculation, begins to become clear. The SAB notes that this terminology is used throughout other interagency manuals such as the Multi-Agency Radiation Survey and Site Investigation Manual, Revision 1 (MARSSIM-Revision 1, U.S. EPA, 2000), the Multi-Agency Radiation Survey and Assessment of Materials and Equipment Manual (MARSSAME, U.S. EPA, 2009); and the Multi-Agency Radiological Laboratory Analytical Protocols Manual (MARLAP, U.S .EPA, 2004), and indeed more widely in environmental protection, which itself provides a very good reason for its adoption. Keeping all those involved in study design or oversight speaking the same language is crucial to the task of providing clarity throughout the site release process.

To address this, there is a need for further introductory material to define and clarify key concepts earlier on. This introductory material will provide a single location in the document where key study design concepts and corresponding MARSSIM terminology is defined, where the rationale for selecting statistical tests and what can be revealed through their usage is explained, as well as where study design and sample size determination can be learned or reinforced for the non-specialists in statistics.

Here the SAB outlines what that material could look like, assuming a target audience of those familiar with basic statistical concepts, specifically, the ideas of null hypothesis and alternative hypothesis, tests of the null hypothesis, Type I error, and power. With this basic framework assumed, the MARSSIM terminology starts with the definitions of "Scenario A" and "Scenario B." The SAB recognizes that MARSSIM has many users who are not well-trained statistically, and the basic statistical concepts may themselves be unfamiliar. For these readers further explication, starting with the most basic statistical terms, will be required. The SAB recognizes also that all the material described below does indeed appear in the MARSSIM document, but the SAB still sees a need for an overall summary to be added early in the document in one location for readers to look for explanations of key study design concepts and the MARSSIM terminology.

Outline

Key Statistical Concepts Review: Introduce basic concepts of study design, namely hypotheses and hypothesis testing, Type I error, power, and introduce several elementary test statistics, including both parametric (e.g. t-tests), and nonparametric (e.g. the Wilcoxon Rank Sum (WRS) test).

Scenario A and Scenario B: These two concepts, which are used throughout the MARSSIM document, relate the null hypothesis to the derived concentration guideline level for average concentrations over wide area ($DCGL_W$). If the null hypothesis is that the average or mean concentration is above the $DCGL_W$ and the alternative hypothesis is that mean concentration is lower than this value, then this is called Scenario A. On the other hand, if these switch, so that concentration specified by the null is less than that specified by the alternative, then this is termed Scenario B. Although Scenario A is the traditional cleanup approach, Scenario B is an alternative available option for cases where the residual radionuclide concentration is indistinguishable from zero or the radionuclide of interest is natural, and its concentration has reached the point of being indistinguishable from natural heterogeneity.

Next should be a discussion of the meaning of rejecting and failing to reject the null hypotheses for each of Scenario A and B. For Scenario A: (1) rejecting the null hypothesis implies that the true (mean) concentration level is very likely to be below the $DCGL_W$ and supports the remediation action to come to a conclusion; and (2) failing to reject the null hypothesis implies that the evidence is consistent with mean concentrations above the $DCGL_W$ and a continuation of remedial action. For Scenario B: (3) rejecting the null hypothesis indicates that the mean concentrations are likely to be above the $DCGL_W$ and a continuation of remedial action is needed; and (4) failing to reject the null indicates that the mean concentrations are consistent with mean concentrations less than the $DCGL_W$ and supports the remediation action to come to a conclusion. The difference is most notable when a study is poorly designed, i.e., it does not have sufficient sample size to meet the power expectations and fails to reject the null more often than desired. Under scenario A, failing to reject the null indicates that the data are consistent with the mean concentration being above the $DCGL_W$. Under Scenario B, failing to reject the null means that the data are consistent with the mean concentration being below the $DCGL_W$ (or action level, AL). Thus, under Scenario B (but not A), a retrospective power analysis must be performed to prove the survey has sufficient statistical power to detect a survey unit that should not have passed.

The Gray Region: In either Scenario A or B, the region of concentrations between that specified by the null and alternative hypotheses is called the gray region. The upper bound of the gray region, or UBGR, is the larger of the two concentrations, and the lower bound, LBGR, is the smaller. The difference between the upper and lower bound, Δ, is termed the width of the gray region. This is often given in units of (theoretical) standard deviations, σ, of the measurements, from the population being sampled, and is termed the relative shift (Δ/σ).

Power, Sample Size, Relative Shift: Specifying appropriate values for Δ/σ, Type I error rate α, and statistical power $1-\beta$ is part of the Data Quality Objectives (DQO) process, and together these determine the necessary sample size. For a fixed relative shift, statistical power using an appropriate statistical test, e.g., the Wilcoxon Rank Sum (WRS) test, increases with sample size. As the relative shift decreases, the necessary sample size increases. If the relative shift approaches zero, then the necessary sample size approaches infinity, thereby making the study design (number of measurements needed) prohibitively expensive. This could be shown visually in MARSSIM Revision 2 with separate figures showing (1) power increases as sample size increases, and (2) necessary sample size increases when relative shift decreases.

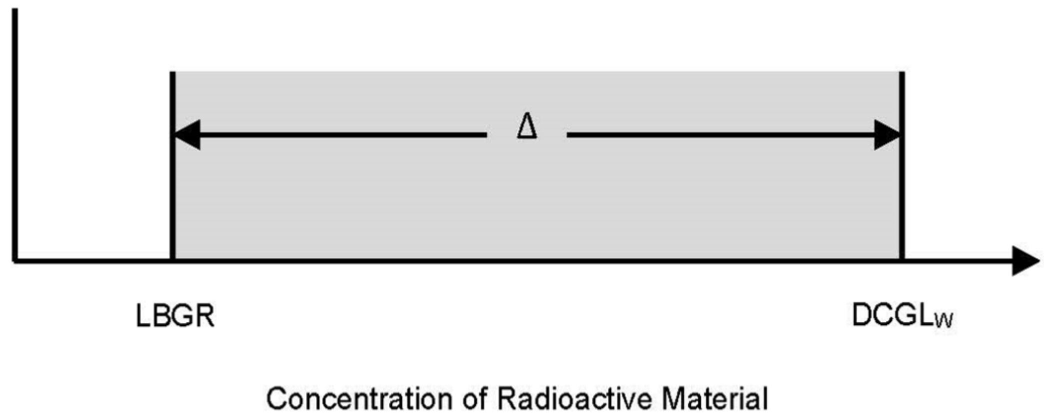

Figure 5.7: Gray Region for Scenario A

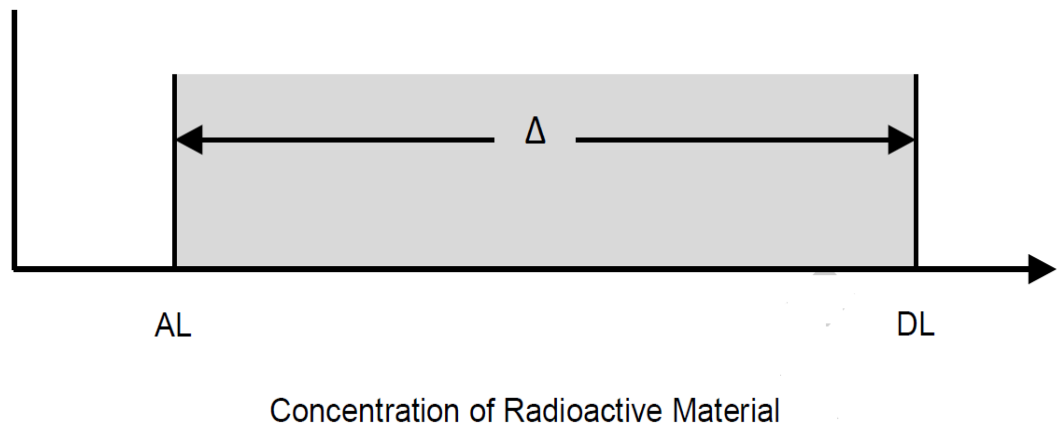

Figure 5.8: Gray Region for Scenario B

In Scenario A, the UBGR is set to the $DCGL_W$ [above see Figure 5.7 of the draft MARSSIM Revision 2 (U.S. EPA, 2020)] and it is suggested that the LBGR be set to a conservative (high-sided) estimate of the amount of residual radioactivity. This requires some knowledge (necessarily uncertain) about the true concentration levels of the site, and how they differ from baseline measurements, as well as the variability of the measurements of concentrations that will be encountered during the actual study. In some cases, this information is available from earlier surveys; in other cases, expert opinion may be all that can be relied upon (see response to Charge Question 3.1). A problem with using Scenario A arises when the $DCGL_W$ is set to nearly zero, implying that essentially no increased concentration is permitted. In this case, the necessary sample size for collection and analysis can become unreasonably large and infeasible, since the relative shift approaches zero. Setting $DCGL_W$ to nearly zero typically implies very small doses and risk from the contaminant, but it could be noted more clearly in the guidance that simply switching from Scenario A to Scenario B doesn't solve this problem, the gray region must be expanded before sample sizes become feasible. The concepts of discrimination limit (DL) and action level

(AL) also need explanation as these are the UBGR and the LBGR respectively under Scenario B [above see Figure 5.8 of the draft MARSSIM Manual (U.S. EPA, 2020)]. The requirement (in order to avoid an "owner-friendly" bias) for retrospective power analysis when the null hypothesis is not rejected under Scenario B should be introduced to provide additional support that successful remediation was based on a robust and defendable decision-making process.

While there is much more statistical content in the draft MARSSIM (U.S. EPA, 2020), these points seem to be the most important for understanding the study-design-related components of the manual.

2.1.2.1 Charge Question 1.2a. *Please comment on the inclusion and proposed implementation of Scenario B (Chapter 4, Section 5.3, and Chapter 8).*

Overall, the proposal for including Scenario B as a recommended approach when the $DCGL_W$ is close to zero (no added contamination) and Scenario A is not feasible was accepted by the SAB. While it is possible to reformulate Scenario A to deal with low $DCGL_W$ by replacing the UBGR with a discrimination limit (DL), it is more natural to revert to Scenario B as long as statistical power is well controlled in the analysis.

2.1.2.2 Charge Question 1.2b. *Is it appropriate to recommend that Scenario B be used only for those situations where Scenario A is not feasible?*

The SAB finds that recommendation of Scenario A as the default approach for designing a survey study is reasonable and agrees that it is also reasonable to only recommend Scenario B when the criteria being tested against are equivalent to there being no residual radioactivity left after remediation. In some situations where the proposed residual radioactive material criterion is "close to zero," consideration of Scenario B may be driven by the available field instrumentation and the laboratory analyses to be performed. Specification of adequate measurement methods is part of the study design. In other situations, consideration of Scenario B may be driven by the fact that the contaminant is also found in the background with high variability in the location to be surveyed. In this case, refinement of instrumentation and analyses will not eliminate the need to rely on Scenario B for ultimate release.

An inadequately designed study should not be used to accept that mean concentration is below a certain level (the $DCGL_W$), which justifies the choice of Scenario A. However, if the concentration level is too close to zero, then the relative shift (Δ/σ) will be too small to give a reasonable number of samples for collection and analysis. Just switching Scenarios from A to B is not sufficient, because the UBGR should be reliably distinguished from the very small $DCGL_W$. A very small $DCGL_W$ can correspond to very small doses and risk, so the basis for the $DCGL_W$ and the benefits of further remediation should be consistent with regulatory requirements. In Scenario B the LBGR would remain small, and renamed as the AL, while the UBGR is set to the discrimination limit (DL) for the lowest concentration that can be distinguished from the AL. If UBGR is enlarged by increasing the $DCGL_W$ so that this increase in the UBGR under Scenario A equals the DL under Scenario B, statistical power may not change with a relative shift that remains the same, except that Type I and Type II errors switch.

A practical issue arises with Scenario B when p-values are evaluated using the estimates of standard deviation computed from the observed data because Type II errors are defined during the design phase. If the relative shift is overestimated in designing the study, p-values would remain valid, but overestimated Type II errors can be expected, which also implies that statistical power is overestimated. For these reasons, it is crucial that Type II errors are properly controlled for in Scenario B to confirm evidence in favor of the null hypothesis (i.e., mean concentration is below the DL). The SAB recommends a retrospective power analysis when using Scenario B to ensure that the survey has adequate statistical power to detect a survey unit that should not have passed.

Language indicating that Scenario B is not as desirable as Scenario A should be dropped or recast. As described, Scenario B can be necessary and important while the burden of proof remains on the MARSSIM user to show (i.e., through retrospective power analysis) that the data strongly support the null hypothesis when Scenario B fails to reject the null hypothesis.

2.1.2.3 Charge Question 1.2c. *Are methods for considering background variability in assessing whether the site is indistinguishable from background reasonable and technically accurate?*

Example 9 in Chapter 8 of the draft manual presents methods that can be used to determine indistinguishability of the survey unit from background when Scenario B is deemed appropriate to use. The recommended approach to adjusting for background variability involves 3 steps. The first is to (1) test for homogeneity in the background levels from reference site to reference site using the Kruskal-Wallis test which is a non-parametric method, and if the null hypothesis of homogeneity is rejected; then (2) estimate a between reference site variance term using a parametric random effects analysis of variance (ANOVA) procedure. Finally (3) the background rate is accounted for by setting the LBGR (in Scenario B) to three times the square root of the estimated between reference site variance ($\widehat{\omega}^2$) component (i.e. LBGR = $3\widehat{\omega}$). The SAB notes that both the Kruskal-Wallis and the random effects ANOVA are very well-established methods, with a caveat being that in testing for significance of the estimated variance components, the ANOVA analysis, using the parametric F test, is sensitive to the normality assumption. Therefore, the SAB agrees on reliance on the Kruskal-Wallis procedure for testing purposes, and only proceeding to step (2) if the test is significant is a reasonable approach.

The SAB notes that the draft MARRSIM document (U.S. EPA, 2020) correctly points out that ANOVA generally provides (in all but extreme cases) reasonable estimates of the between reference site variance component, even though the p-values from the test may be unreliable; in other words, the results of the ANOVA procedure are appropriate for estimation of variance components even when the F test is not reliable for testing the significance of those components if there is failure of the normality assumption.

The SAB also notes that the third step in the adjustment for background radiation seemed to be overly forgiving with the question being: why are 3 standard deviations of the random effect chosen as the adjusted LBGR? This is justified in NUREG-1505 (NRC, 1998) by looking at the probability (Pr) that an arbitrarily distributed random variable (X) is more than t standard deviations (σ) away from its mean (μ) using Chebyshev's inequality:

$$\Pr(|(X-\mu)| \geq t\sigma) \leq 1/t\text{^}2.$$

With t equal to 3, the probability is less than 1/9 = 0.11. (Actually a one-sided bound is more relevant to this problem, and is given as (Prob(X - $\mu \geq t\sigma$) < $1/(1+t^2)$) or 0.10 when t = 3). However, this bound can be very weak, for a normally distributed random variable the probability of exceeding this value is just 0.0013. The SAB recommends that t =1.3 (upper tail probability 0.90 for normal distribution) be considered. Here σ is the standard deviation of the between reference site random effect estimated in the ANOVA (in Example 9 of Chapter 8 of the draft MARSSIM document (U.S. EPA, 2020), it is termed omega, $\widehat{\omega}$).

Regarding clarity of presentation, the SAB agrees that the use of the Kruskal-Wallis test and the use of random effects ANOVA to estimate the between reference site component were adequately described in Example 9, Chapter 8 of the draft manual. However, the presentation is challenging and can be clearer. For example, Example 9 is built on Equations 13-3 and 13-13 as well as tables (e.g. Tables 13.1 and 13.5) from NUREG-1505 (NRC, 1998) without providing details about the equations and tables in the text. There is also a typo in the calculation of $\widehat{\omega}$ in Example 9 (i.e., $\widehat{\omega}$ should be calculated from $\sqrt{0.55}$, not from $\sqrt{55}$).

2.1.2.4 Charge Question 1.2d. *Is the inclusion and proposed implementation of added requirements for retrospective power analysis and the Quantile Test while using Scenario B technically appropriate and discussed adequately and clearly?*

The draft MARSSIM document adopts statistical methods that test for both changes in location (with emphasis on shifts in the median) and changes in the upper portion of the distribution; both are important for the evaluation of residual radioactivity. The main issues in the analysis of data under Scenario B are: (1) whether appropriate statistical tests have been used for data analysis and (2) whether sample size and power had been appropriately calculated in the design phase. Regarding point (2), the SAB agrees that retrospective power analysis is fundamental to acceptance of null hypothesis results when Scenario B is utilized. The key to retrospective power analysis is comparing the observed relative shift ($\Delta/\hat{\sigma}$) using the value, $\hat{\sigma}$, the standard deviation estimated during the study, to the Δ/σ used in the design phase, as further described in Appendix M of the draft manual.

Regarding point (1), the recommended, nonparametric, tests under Scenario B, are the Wilcoxon rank sum (WRS) test and Quantile test. Down on the list are parametric tests (one and two sample t-tests). An argument may be raised against nonparametric tests in general, since they mostly are used to reduce the influence of outliers (e.g., skewness and other failures to be normally distributed) on the estimation of location parameters. Both the Sign and the WRS test are testing for a shift in the median rather than the mean and as such are very robust to the influence of outliers. However, possibly because risk models are generally linear in radiation exposure, the focus in MARSSIM, is stated to be on controlling mean exposure. Because of this, a combination of the WRS test and the Quantile test is recommended in Scenario B, as well as inspection of the sample mean compared to the median. The Quantile test is sensitive to shifts that only affect some of the data (i.e., the upper portion). Furthermore, values that exceed an investigational limit are to be examined more closely, despite the outcome of the non-parametric tests (see Section 8.6.1 of the draft manual).

The description of the use of the WRS test is clear and can be readily implemented in spreadsheet format (and almost all general-use statistical packages provide this test). Section 8.4.2 indicates that to apply the WRS test for Scenario B, first the AL (here equal to zero) is subtracted from the survey unit measurements, and then what boils down to a one-sided WRS test is performed on the adjusted measurements. Otherwise, the Quantile test is performed (again on the adjusted survey unit measurements), and if this test also is not rejected then the survey unit is judged as not exceeding the release criteria. Table 5-2 is provided for power calculations for the WRS test based on Equation O-1 and can be applied both prospectively and retrospectively. No power calculations are given for the quantile test. This is a problem for Scenario B since retrospective power calculations are considered essential before accepting null results. It is recognized that the same sample size will be used for both tests, but it seems possible and desirable to put more investigation into the power characteristics of the quantile test so that power for specific scenarios can be determined either theoretically or by use of a simulation study, at least in simple cases. For example, if the upper decile of the true population is shifted upwards high enough so that the mean is above the $DCGL_W$, but the median is not, then it seems possible to compute power for the quantile test according to sample size. Similar power calculations can be performed if it is the upper 5 percent, or 1 percent that is so shifted.

In the draft MARSSIM document, the intent of the Quantile test seems to guard only against the worst possible scenarios involving non-normal tail distributions in the reference or survey areas (or both). Another possible concern is that if the quantile test is very low power, then giving it half of the Type I error will weaken the power of the overall WRS + quantile test. However, given difficulties in testing for differences in the means for skewed data, the combination of a Wilcoxon and Quantile test appears to be a reasonable choice.

Overall Recommendations for Charge Question 1.2:

- The SAB supports the inclusion of Scenario B.
- The SAB agrees that it is reasonable to only recommend Scenario B when Scenario A is infeasible (i.e., when the $DCGL_W$ is close to zero).
- The SAB notes that "close to zero" is a concept that may be dependent on instrumentation and that defining the appropriate instrumentation is also a key aspect of study design.
- The SAB requests that language indicating that Scenario B was not as desirable as Scenario A be dropped or modified. It should be emphasized that Scenario B has an important and necessary role to play.
- The presumption that the burden of proof shifts from user to regulator when Scenario B is utilized should be retracted. The burden of proof remains on the MARSSIM user to show (i.e., through retrospective power analysis) that the data strongly support the null hypothesis when Scenario B fails to reject the null.
- The methods for considering background variability in assessing whether the site is indistinguishable from background are technically accurate. The use of $3\hat{\omega}$ as the LBGR in

Scenario B (Example 9) is overly permissive in the SAB's estimation, with little chance for the survey unit failing the null hypothesis that the survey unit concentrations do not exceed the LBGR. The SAB recommends consideration of somewhat smaller values as more in keeping with the generally conservative approach of MARSSIM.

- The SAB agrees that the inclusion and proposed implementation of added requirements for retrospective power analysis under Scenario B are appropriate. The document would benefit by having more examples illustrating the calculation of retrospective power. Too few worked examples are currently given.

- The discussion of the Quantile Test while using Scenario B is technically appropriate. However, a better description of the Quantile Test, literature references, discussion of the underlying null and alternative hypotheses, and power, are all needed in the MARSSIM revision.

- Clarity throughout the document could be improved as well as access to key concepts/references. In some cases, concepts are used before they are defined, or defined differently in different places.

- MARSSIM should acknowledge underlying assumptions of statistical tools and include cautionary notes indicating under what conditions the statistical tests become unreliable.

- More complete step-by-step, stand-alone examples and case studies with all statistical tools worked out in detail would enhance MARSSIM's utility during the site release process.

- The SAB recommends that further introductory material be developed and aimed at an audience consisting of non-specialists in statistics to introduce in one place the statistical concepts of null and alternative hypothesis, tests (e.g., WRS, Sign, t, ANOVA, etc.), and the terminology used throughout, including Type I errors, Type II errors, sample size, power, the gray region, σ and Δ, and the AL, DL, UBGR, and LBGR, under both Scenarios A and B.

2.1.3. Charge Question 1.3. Measurement Quality Objectives

2.1.3.1 Charge Question 1.3a. *Is the proposed implementation of the concept of Measurement Quality Objectives adequately and correctly described, including the concept of measurement uncertainty (Chapter 4 and Appendix D)?*

Measurement Quality Objectives (MQO)s define performance requirements and objectives in the measurement system. In MARSSIM Revision 2 draft (U.S. EPA, 2020), MQOs that are proposed for consideration (in Chapter 2, 4, 6 and Appendix D) include the following:

- the method's uncertainty at a specified concentration, usually at the UBGR (expressed as a standard deviation)

- the method's detection capability (expressed as the minimum detectable concentration, or MDC)
- the method's quantification capability (expressed as the minimum quantifiable concentration, or MQC)
- the method's range, which defines the method's ability to measure the radionuclide of concern over some specified range of concentrations
- the method's specificity, which refers to the ability of the method to measure the radionuclide of concern in the presence of interferences
- the method's ruggedness, which refers to the relative stability of method performance for small variations in method parameter values

The SAB focused its review on the three most important MQOs (method uncertainty and related measurement uncertainty, detection capability, and quantification capability) in the draft MARSSIM Revision 2 document. The SAB provides the following comments regarding these parameters:

<u>Method Uncertainty and Related Measurement Uncertainty:</u>

The SAB finds the description of the concepts of method uncertainty and measurement uncertainty and their implementation to be inadequate and unclear.

The development of MQOs for a project depends on the selection of an action level (usually the UBGR) and gray region (see response to charge question 1.2) for each measure and (analyte) during the survey planning process. The gray region is a set of concentrations between the action level and a project determined lower discrimination limit where the project planning team is willing to tolerate a specific error rate. There are three concepts that need much more specific definition for this MQO:

First, method uncertainty is the *a priori* uncertainty of a specific method. It is the predicted uncertainty of a measured value that would be calculated if the method were applied to a hypothetical sample with a specified concentration, typically the UBGR. Reasonable values for measurement method uncertainty can be predicted for a particular measurement technique based on typical values for specific parameters (e.g., count time, efficiency) and previous surveys of the areas being investigated.

The term "measurement method uncertainty" is used in the draft MARSSIM document. Although MARLAP (U.S. EPA, 2004) uses the term "Method Uncertainty," the SAB finds the term "measurement method uncertainty" acceptable.

Second, the concept of Required Method Uncertainty should be better defined in the draft MARSSIM document. It is an *a priori* upper limit for the method uncertainty, at and below the UBGR, to ensure that the selected measurement method can reliably perform measurements at the most critical concentration level for the survey. This *a priori* MQO should be based on the data quality objectives (or project requirements) of a tolerable error rate and width of the gray region in the decision-making process. Identifying a required method uncertainty helps to select which instrument(s) will be able to perform the project specific measurements. While Section 2.3.1 states:

> "the required measurement method uncertainty is calculated based on the width of the gray region and is related to the minimum detectable concentration (MDC),"

there is no mention of how to determine the *required measurement method uncertainty,* or how it should be used in planning of the survey in Chapter 4, Chapter 6, or Appendix D. This deficiency should be addressed in the MARSSIM Revision 2 document. In Appendix C of MARLAP (U.S. EPA, 2004) for laboratory analysis, it is recommended that if decisions are to be made about the mean of a sampled population, the required method uncertainty (u_{MR}) be less than or equal to the width of the gray region (Δ) divided by 10 for sample concentrations at the upper bound of the gray region (typically the action level). If this method uncertainty cannot be achieved, then an uncertainty as large as $\Delta/3$ may be allowed if σ_S is small or if more samples are taken per survey unit. This approach may be considered for the MARSSIM document. Additionally, the exact means of calculating a *required method uncertainty* in a general formula is needed, as well as the range over which this MQO is in force.

The Required Method Uncertainty is also used to evaluate if the measurement uncertainty (discussed below) calculated from measured values for specific analytes from the site survey meets the project MQO. This use of the Required Method Uncertainty should be clearly stated in the MARSSIM document.

<u>Measurement Uncertainty</u>:

Third, *measurement uncertainty*, is an *a posteriori* calculated value. The measurement uncertainty is the uncertainty of <u>all</u> the measurements made for an analysis on a specific sample; it collects all *a posteriori* measurement uncertainty data to determine the expanded uncertainty for the sample measurement. This also should be called the combined standard uncertainty (CSU) or the combined uncertainty with a noted coverage factor, k[2]. This means that the probability of exceeding the UBGR is the same for all measurements (ensuring the specified Type I error is the same for the analytical results in the gray region).

In the discussion of measurement uncertainty in Section 6.4, the SAB finds the draft MARSSIM document does not identify the different components of measurement uncertainty and does not conform to the terminology provided in NIST Technical Note 1297 (NIST, 1994) and the ISO Guide to the Expression of Uncertainty in Measurement (ISO, 2015b), or references cited in the MARSSIM draft document. The MARSSIM document must include clear statements about measurement uncertainty, including at a minimum, the components of uncertainty, the means by

---

[2] From MARLAP, Introduction, p 1-8, "The combined standard uncertainty may be multiplied by a specified factor called a coverage factor (e.g., 2 or 3) to obtain an expanded uncertainty (a two-sigma or three-sigma uncertainty), which describes an interval about the result that can be expected to contain the true value with a specified high probability."

which the parameters of uncertainty (e.g., standard deviation) have been determined and the approach used to combine the multiple uncertainties into a combined and expanded uncertainty. Specifically, descriptions of the components of uncertainty that should be included are:

- those from sampling design
- any of the measurement processes used to obtain final results
- uncertainty in sample collection[3]
- identifying the measurement model(s) (i.e., equation used to convert measurement components into radioactivity units - this is the starting point for developing the list of uncertainty components, after which additional Type B components are also included)
- requirements for documenting how the uncertainties were determined and reported [particularly when non-traditional methods are used]

The description should provide examples of the known influences and whether the resultant uncertainty from these influences is Type A (derived from statistical methods) or Type B (usually derived by non-statistical means such as experience, manufacturer's specifications, reference data, reported calibrations/reports, or expert knowledge) and the specific means by which their quantitative parameters (estimates of the standard deviation) were derived.

Section D.4.2.4 in Appendix D lists sources of uncertainty for inclusion in uncertainty determinations.

> "The uncertainty of a measurement expressed as combined standard uncertainty includes the counting uncertainty of the measurement instrumentation and the sum of the errors associated with the measurement system… Uncertainty factors associated with the measurement system for scanning and direct measurements can include variability in the distance between the detector surface and the sampling media, variability in the speed at which a detector passes over a survey point (or the amount of time the detector is held over the sampling point for direct measurements), the extent to which interference from other radioactive sources is minimized, and the extent to which human performance factors create variability in the measurement system. Uncertainty factors associated with sampling include variability in the sample collection methods and variability in the distribution of residual radioactive material in the sampling media".

The list provides good examples of sources of uncertainty and should be cross-referenced in the main text or add some of the examples in the description. However, the list does not elucidate the Type A and Type B components of uncertainty, or how these affect the combined standard uncertainty of a measurement process and the MQOs.

There is no discussion of whether or not correlated measurement uncertainties[3] are part of the MQO process. If the covariance of correlated uncertainties is non-trivial, plans to mitigate their effects (e.g., by calibration to laboratory readings and controlling for operator differences in the design of the study) should be included as part of the discussion in Chapter 6 or Appendix D.

---

[3] The NIST Uncertainty Machine (http://uncertainty.nist.gov) can assist in these calculations.

Equation 6-18 captures only the uncorrelated components of the uncertainty. However, Equation 19-11 in MARLAP (U.S. EPA, 2004) is more robust and contains both the uncorrelated and correlated components. To avoid ambiguity, it is suggested that the 19-11 equation that includes the correlated uncertainty components be referenced so that the user will be able to include correlated measurement uncertainties when they exist and keep the calculations defensible.

Detection Capability:

The SAB finds the description of the concept of detection capability and its implementation in the draft MARSSIM document to be generally adequate and correctly described.

Section 2.3.1 of the draft MARSSIM document describes the minimum detectable concentration (MDC) as the MQO for defining the detection capability of the measurement system.

The MDC is the *a priori* activity concentration that a specific instrument and technique can detect with a specified probability (typically 95 percent) of producing a net count (or count rate) above the critical level (L_C). The detection limit (L_D) is the net response level that can be expected to be seen with a detector with a fixed level of confidence. The MDC is the detection limit multiplied by an appropriate conversion factor to give units of activity.

The MDC for a specific instrument can be calculated using Equation 6-3 through 6-5 in Chapter 6. These equations provide the basics of using the measured instrument background count, B, and Type I and Type II error rates to determine the critical level and detection level in terms of counts only.

On Page 6-9, lines 2-13, the MARSSIM equation for the Critical Level [Lc], Detection Limit [L_D] and MDC are based on the determination of B, the mean background counts [defined on MARSSIM, p. 6-8, line 11].

L_C = Critical Level $= L_C = 2.33\sqrt{B}$ (6-3)

L_D = Detection Limit $= 3 + 4.65\sqrt{B}$ (6-4)

$= 3 + 2 * L_C$

MDC $= C * L_D = C * [3 + 2 * L_C]$

$= C * [3 + 2 * 2.33 * \sqrt{B}]$, (6-5)

A panel member suggested there may be an error in the notation for “B” in Equation 6-1 through 6-5 in Chapter 6. The panel member commented that field blanks are usually collected and used in the laboratory to assess contamination associated with sampling, transport and laboratory procedure, and suggest “B” in Equation 6-1 to 6-5 should be denoted as blank counts. However, Chapter 6 is on field measurement methods and instruments. In Lloyd Currie’s seminal paper (Currie, 1968), B is the net signal which is identical, in principle, to the sample of interest (or for a scanning measurement the area of interest) except that the substance to be quantified is “absent” and should not be confused with laboratory quality control blanks (e.g., method blank, reagent blank). While the follow-up paper to Currie (1968) by Brodsky (1992) used “blank” and "background" interchangeably, and denoting B as blank is not incorrect, it is probably clearer to denote B as the average instrument background count, as was used in the draft MARSSIM document. On Page 6-9, lines 2-13 the description of Lc, L_D, and MDC are consistent with NUREG-1507 (NRC, 2020)

It should be noted that Equation 6-5 only applies to detection systems that operate in pulse or counting mode, it is not applicable to current mode detection systems, for example an ion chamber or thermoluminescent dosimeter (TLD). The MDC corresponds to the smallest activity concentration measurement that is practically achievable with a given instrument and type of measurement procedure. As noted in NUREG 1507 (NRC, 2020), unlike the detection limit (L_D), which may be a count or count rate, and is independent of field conditions, MDC depends not only on the particular instrument characteristics (instrument efficiency, background, integration time, etc.), but also on the factors involved in the survey measurement process (U.S. EPA, 1980), which include surface type, source-to-detector geometry, and source efficiency. These concepts should be explained more clearly in the MARSSIM Revision 2 document.

Moreover, an MDC should be evaluated with all known sources of uncertainty being properly quantified and combined into an appropriate expanded uncertainty. Equation 6-5 (Section 6.3.1) treats the calibration factor as if it had no uncertainty and disregards other uncertainty sources that arise from adjusting a laboratory calibration to a field calibration. The example (Example 1) is too rudimentary and omits the need to consider more than just counting uncertainties. A more complete example could be presented in Appendix O. A similar situation exists in the discussion of surveyor efficiency in Section 6.3.2.1 (page 6-17). The example (Example 5) states that surveyor efficiency can range between 0.5 and 0.75. The MARSSIM guidance suggests the use of a constant of 0.5 as a matter of conservatism with no uncertainty consideration. If the NIST and ISO measurement uncertainty approach had been implemented, then the surveyor efficiency would be better expressed as 0.625 with an uncertainty of 0.1 or 0.14 assuming a Type B uncertainty with a triangular or rectangular distribution respectively. MARLAP (U.S. EPA, 2004) suggests that the lower 95 percentile of the distribution of uncertainty in the calibration factor be taken as a conservative value for the calibration factor. The SAB recommends that further development of examples illustrating these approaches be considered.

The current text of MARSSIM needs to be changed to reflect the importance of selecting a method uncertainty first before making measurements. Values of MDC or MDA (or MQC) are not detection or quantification values. The iterative nature of a MARSSIM-described project should improve what potentially can be determined. Only an *a posteriori* determination based on critical level (or decision level) and measurement uncertainty or calculation of a confidence interval should be used to determine detection. Via this iterative process, uncertainty characterization and method selection improve over the stages of the project with FSS values of uncertainty better characterized than those available at the time of scoping or site characterization.

The manual should give due consideration regarding the overall derivation of MDC or MDA because the manual references many measurement methods that do not follow the normal counting statistics used in the rudimentary examples given in Chapter 6. Such methods include the dose integrating technologies of TLD, optically stimulated luminescence (OSL) and electrets as well as other methods listed in Appendix H such as, x-ray fluorescence analysis, mass spectrometry and phosphorescence analysis by laser.

Regarding setting the required MDC for a specific analyte for selection of instrumentation during the project planning phase, Section 2.3.1 (pg 2-13 of the draft MARSSIM document) recommends that the MDC should be less than 50 percent of the $DCGL_w$ in Scenario A and the DL in Scenario B. (It should be noted that the LBGR is usually set at 50% of the $DCGL_w$ in Scenario A, see response to charge question 3.1). The SAB agrees with this recommendation but would like to clarify the concept with this sentence: "For both Scenario A and B, the MDC should be less than ½ the UBGR. For Scenario A, the UBGR is the $DCGL_W$; while for Scenario B, the UBGR is termed the discrimination limit (DL)." When the MDC reported for a radionuclide is near the DCGL, the confidence in quantitation may be low. In essence, the required project MDC for a specific radionuclide is set at 50% of the UBGR. Appendix D, Section D.4.2.4 of the draft MARSSIM document also recommends that "If the radionuclide result is below the MDC, report the actual result of the analysis. Do not report data as "less than the detection limit." Even negative results and results with large uncertainties can be used in the statistical tests described in Chapter 8." The SAB also agrees with these statements.

Minimum Quantifiable Concentration (MQC):

In the Multi-Agency Radiological Laboratory Analytical Protocols Manual (MARLAP) (U.S. EPA, 2004), the minimum quantifiable concentration, or the minimum quantifiable value of the analyte concentration, is defined as the concentration of analyte in a laboratory sample at which the measurement process gives results with a specified relative standard deviation. A relative standard deviation (or coefficient of variation[4]) of 10 % is usually specified. Unlike MARLAP (USEPA, 2004) and MARSAME (U.S. EPA, 2009), which include the concept of the method's quantification capability (expressed as the minimum quantifiable concentration, or MQC), MARSSIM takes a different approach by incorporating requirements for quantification capability into detection capability and recommending that the MDC be less than the 50 percent of the UBGR. MQC is only listed in the draft MARSSIM document (U.S. EPA, 2020) on the list of MQOs, but the document has no discussion of how it is determined (it is listed in Section 4.8.2 but not in D.1.7.1). The question of whether measurement quantifiability should be incorporated further into MARSSIM, Revision 2 is discussed below in response to charge question 1.3e.

In addition, the SAB finds that in the case of each MQO, there is no guidance in the draft document on what to do if a sample measurement does not achieve a required MQO. In MARLAP, when MQOs are not achieved during analysis, the data are qualified in the data review process so that the client is aware that a project MQO has not been achieved.

2.1.3.2 Charge Question 1.3b. *Is the proposed calculation of measurement uncertainty consistent with the concept of Measurement Quality Objectives?*

The SAB found that this charge question was a little unclear but tried to answer the question of whether the calculations of measurement uncertainty were reasonable and consistent over the document. The general approach towards the calculation of measurement uncertainty is based

[4] Coefficient of variation is the ratio of standard deviation over the mean

upon the assumption that counting statistics follow the Poisson distribution, with mean equal to the variance of such a random quantity. Random errors that follow the Poisson distribution are then propagated (e.g., re-expressed and/or combined with other random quantities) using equation 6-18 (the error propagation equation). The assumption that the counts follow a Poisson generally is true in theory, although in practice counting efficiencies need to be factored into the counting since measurements cannot capture every disintegration.

It is evident that the main focus of the discussion in Chapter 6 is on testing whether a sample of potentially contaminated material has significantly higher activity (at a given value of $z_{1-\alpha}$) than a background sample. Statistical inference relevant to MARSSIM however is not restricted to testing of this null hypothesis. As is made clear in other sections of MARSSIM, testing whether a given site is below a DCGL is crucial. One important difference in calculations is that a properly designed comparison between contaminated and background samples removes the effect of systematic errors, if they are common to both sample and background since they are subtracted out in the comparison, this subtraction would not occur when testing for compliance with a DCGL. Section 6.4 (Measurement Uncertainty) discusses systematic uncertainties (errors in counting efficiencies, activity measurements, calibration factors) but concludes with statements to the effect that by judicious laboratory practice, training, SOPs, etc. systematic errors should be minimized. No attempt to include systematic uncertainty into formal calculations of uncertainties is attempted. Clearly limiting systematic uncertainties by good laboratory and field operating procedures is crucial. Some attempt to include them into the error propagation is nevertheless appropriate and should be included as part of the document.

Other issues related to systematic uncertainties arise when considering scanning vs. sampling for site investigation. Laboratory analysis of samples taken from the site may be much more expensive than scanning. In MARSSIM-Revision 1 (U.S. EPA, 2000), it was stated that random sampling uncertainties would outweigh systematic uncertainties because the total fraction of the site monitored by sampling was very low. On the other hand, scanning could cover as much as 100 percent of an investigational area (site), recording perhaps thousands of counts. In this case it could be argued that the impact of random errors would disappear while systematic errors, even if no larger than those inherent in laboratory analysis, would become the prominent source of measurement uncertainty. The issue of whether systematic uncertainties are greater for scan only surveys than for laboratory readings is not well-addressed in the draft MARSSIM document. There is, in Chapter 6, Section 6.3.2, an extensive discussion of scanning minimum detectable count rates (MDCR) for a two-stage scanning approach in which an operator scans a large region quickly but pauses to measure longer when a possible increase in count rate is detected. However, whether readings conducted in this way are comparable in precision to laboratory (or direct) measurements is not discussed. It is stated (page 6-34) that FSS can be conducted using a scan only design to demonstrate that a site is not above a DCGL. It is stated on page 5-43 Section 5.3.6.1:

> "The scan-only methodology will require validation, which likely requires collecting some percentage of samples for laboratory analysis to compare with results from the same location."

The statistical discussion of uncertainties includes Equation 6-17 from the MARSSIM document, which again is relevant to comparisons between impacted vs. background since the difference in net count rates will have standard error as given in (6-17). This could usefully have been broken up into parts, namely the measurement uncertainty for the count rate for the background, the uncertainty for the count rate for the sample, and for the difference between the two.

The SAB is of the opinion that the proposed calculation of measurement uncertainty is consistent with the concept of Measurement Quality Objectives (if the discussion is restricted to the discussion of the formula for the standard deviation of the net count rate of the net count σ_n and the total uncertainty). However, as discussed in response to charge question 1.3a, measurement uncertainty calculated from measured values for specific radionuclides from the site survey has to be compared with the required method uncertainty to determine if the project MQOs for the radionuclides are met. This part of the DQO process is missing in the current draft MARSSIM document.

The uncertainty of a measurement expressed as combined standard uncertainty includes the counting uncertainty of the measurement instrumentation and the sum of the errors associated with the measurement system. The draft MARSSIM document provides equations to calculate the total uncertainty associated with the counting process, both the background measurement uncertainty and the sample measurement uncertainty. The standard deviation of the net count rate, or the statistical counting uncertainty, can be calculated using Equation 6-17 (from the draft MARSSIM document) given below:

$$\sigma_n = \sqrt{\frac{C_{s+b}}{t_{s+b}^2} + \frac{C_b}{t_b^2}}$$

where

- σ_n is the standard deviation of the net count rate result
- C_{s+b} is the number of gross counts (sample)
- t_{s+b} is the gross count time
- C_b is the number of background counts
- t_b is the background count time

The standard deviation associated with the total uncertainty can also be calculated. The draft document assumes the individual uncertainties are relatively small, symmetric about zero, and independent of one another, so that the total uncertainty for the final calculated result can be approximated by the following equation:

$$\sigma_y = \sqrt{\left(\frac{\partial y}{\partial x_1}\right)^2 \sigma_{x_1}^2 + \left(\frac{\partial y}{\partial x_2}\right)^2 \sigma_{x_2}^2 + \cdots + \left(\frac{\partial y}{\partial x_n}\right)^2 \sigma_{x_n}^2} \qquad (6\text{-}18)$$

where $y = f(x_1, x_2, \ldots x_n)$ is a formula that defines the calculation of a final result as a function of the collected data. All variables in this equation (i.e., $x_1, x_2, \ldots x_n$) are assumed to have a measurement uncertainty associated with them and do not include numerical constants. σ_y is the standard deviation, or uncertainty, associated with the final result, and $\sigma_{x_1}, \sigma_{x_2}, \ldots \sigma_{x_n}$ are the standard deviations, or uncertainties, associated with the parameters $x_1, x_2, \ldots x_n$, respectively. Equation 6-18 from the draft MARSSIM document, generally known as the error propagation formula, can be evaluated to determine the standard deviation of a final result from calculations involving measurement data and their associated uncertainties.

It is important to note that correlated uncertainties will exist in some cases. A more general error propagation equation allowing for correlated uncertainties is presented as Equation 19-11 in MARLAP (U.S. EPA, 2004) [and as equation A-3 in NBS TN 1297 (NIST, 1994)]. The equations cited in MARLAP and TN 1297 are the same. MARSSIM should acknowledge in the discussion of Equation 6-18 that a more complete formula is required if correlated uncertainties exist.

A description of the standard uncertainty, σ_{x_1}, for each input parameter of the final calculation needs to be stated. Once the individual standard measurement uncertainties for each parameter are introduced, the analysis requires that there be a determination of whether or not any of those parameters are correlated. Once that has been established, the combined standard uncertainty may be calculated using the individual standard uncertainties and the estimated covariances. (Equation 6-18 in Chapter 6 assumes that *none of the standard uncertainties* of the input parameters are correlated.) A reference to how to include covariances on two correlated measurement parameters should be included. Furthermore, it should be stated that the measurement uncertainty is referring to the combined standard uncertainty (this is the one sigma uncertainty that includes all components of uncertainty), or the expanded uncertainty with the coverage factor, k. This is not present in the current draft MARSSIM document.

The description of measurement uncertainties is continued on Page 6-29, lines 8-12, and assumes that uncertainties for correction factors are small, symmetrical about zero and non-correlated. This is seldom the case. The SAB believes that additional text is necessary to indicate to the reader what to do in the case where the individual uncertainties are not small, not symmetric about zero, and not independent of one another. When uncertainties are on the order of 10-20%, the second order effects to the Taylor Expansion must be considered. The effect on Equation 6-18 will be more significant when the uncertainties get larger approaching the MDC (although the uncertainty at the MDC or LBGR should both be less than a *required method uncertainty*. Again, this needs to be identified in this document). Utilizing measurement technology with higher sensitivity may offer a viable option to limit the expansion of Equation 6-18.

On Page 6-31, line 1, Example 8 illustrates the point for computing an expanded uncertainty at a given confidence level. However, the number of background counts is only one part of a measurement equation or the total uncertainty. Inclusion of a more complete example would be much more instructive thus taking the reader through the whole process: measurement equation, additional sources of Type B uncertainties, list of uncertainty components, estimating values for uncertainty components, sensitivity factors, correlated uncertainties, combining the uncertainty components into the standard combined uncertainty, expanding the uncertainty at confidence levels, and reporting the uncertainties. Additionally, it would be advantageous to the reader to move Example 8 into Chapter 4.

The SAB has some concerns with how the material on uncertainty is organized in the manual. Chapter 4, in Section 4.8.2, briefly mentions measurement uncertainty as a possible MQO and refers to Chapter 6 in regard to how to calculate it but does not mention Appendix D at all. Chapter 6, in Section 6.4.2, gives a basic calculation of counting uncertainty and in Section 6.4.3 gives basic propagation of error formulae, but does not give any other examples, and again does not mention Appendix D. Example 8 in Section 6.4.4 shows how to calculate measurement uncertainty as a very simple and uncomplicated case. This is not what will be encountered in a real project where many different sources of uncertainty must be determined. Specifically, the types of uncertainty used should be categorized as:

- Type A uncertainty: "A Type A evaluation of standard uncertainty may be based on any valid statistical method for treating data."
- Type B uncertainty: "A Type B evaluation of standard uncertainty is usually based on scientific judgment using all the relevant information available."[3]

Then their method of determination has to be described in the project documents before including in the final determination of measurement uncertainty. For users of this manual, the only calculation shown would be misinformative and leave them with lack of direction on how to apply uncertainty calculations to more complicated situations that will be encountered.

2.1.3.3 Charge Question 1.3c. *Is the method appropriate and practical for both laboratory and field (including scan) measurements?*

This part of the question deals with the application of calculating measurement uncertainty as it applies to field as well as laboratory measurements. Both laboratory and field measurements including scan-only are discussed in the draft MARSSIM document, with Chapter 6 covering topics relevant to the use of scan-only and other field instrumentation. The computation of minimum detectable concentrations is extended to field measurements, especially measurements based on scan-only instrumentation. While the treatment of scan-only and other field measurements is overall quite strong in Chapter 6, there are some issues that need further attention. For example, for scan-only surveys detectability is emphasized as opposed to estimation, as in the extended discussion around Example 2 and Example 3. Section 6.6.1 explicitly states that (page 6-33) "Scanning is used in surveys to locate radiation anomalies by searching for variations in readings, indicating gross activity levels that may require further investigation or action." This statement, however, is partially contradicted on page 6-34 which in reference to scan-only surveys states:

> "Important considerations include that the scan MDC and measurement method uncertainty are sufficient to meet MQOs to both quantify the average concentration of the radioactive material and to identify areas of elevated activity."

It is unclear on what basis scanning would be precluded as a tool for quantifying the amount of radiation at a site. Presumably this would imply there are systematic errors or uncertainties in the scanning methods that are consistently greater in magnitude than in the laboratory-based methods. If this is the case, then these uncertainties should be listed to support the advice given. Another gap in the treatment arises regarding the contribution to measurement uncertainty of systematic versus random uncertainties. This can be seen in the discussion of Equation 6-5 where random uncertainties in the counts data are carefully characterized, but (systematic) uncertainties in the calibration factor C are only fleetingly described. As increasingly large fractions of the site are measured for the purpose of quantification, the errors in the factor C will become the dominant source of uncertainty in the system, so that characterizing their uncertainties is crucial. It would be helpful if the discussion around Equation 6-5 would consider the limiting case when the total number of readings or count times are so large that random uncertainties in the mean concentration estimate shrink to negligibility, but the error distribution of C remains unchanged. Further description of the sorts of experiments that can be performed to characterize uncertainty in the factor C (which is assumed to affect all measurements equally) may be helpful to those users faced with the problem of characterizing measurement uncertainty in field (especially scan-only) measurements. Detailed discussion of the derivation of detection probabilities for alpha scanning is given in appendix J, but no uncertainties in scanning efficiencies are considered therein. Nothing is said either about quantification rather than detection as a goal of such scanning.

Section 6.4 provides a discussion of measurement uncertainties including discussion of systematic vs. random uncertainties. As noted elsewhere in the SAB review, the advice provided in Section 6.4.1 emphasizes the minimization of systematic uncertainties, e.g., by developing and following standard operating procedures and by selection of suitable instruments for the problem at hand; for example, the section advises beta scanning rather than alpha scanning of a porous concrete surface. Overall, the discussion in Section 6.4 seems helpful for those conducting both laboratory and field measurements, however all derivations of measurement uncertainties refer to random uncertainties. Section 6.4 does not attempt to incorporate systematic error into formal error propagation (e.g., through modifications of the error propagation equation, 6-18). Page 6-28) states that,

> "It is difficult to evaluate the systematic uncertainty for a measurement process, but bounds should always be estimated and made small compared to the random uncertainty, if possible. If no other information on systematic uncertainty is available, Currie (NRC, 1984) recommends using 16 percent as an estimate for systematic uncertainties (1 percent for blanks, 5 percent for baseline, and 10 percent for calibration factors)."

The advice given by Currie does not seem relevant to field measurements and may, for laboratory measurements, be overly optimistic or aspirational rather than constituting worthwhile expert advice. It is important to use whatever information is available to back up generic statements as these, or perhaps rephrase them to indicate that assigning 16 percent or other estimates to total systematic uncertainty should be technically defensible. Section 6.6.5 describes instrument calibration in some detail. Factors that affect calibration validity are also described as well for a variety of field measurements.

Overall, the SAB finds that the methods for characterizing uncertainty in measurements described in the draft MARSSIM document are generally valid for both laboratory and field measurements. The two suggestions provided in reference to this charge question are: (1) to describe in more detail the role of scan-only methods in quantification of mean concentration, as well as in detection of anomalies, and (2) to further develop examples to include systematic uncertainties in error propagation, i.e., to include systematic uncertainties in the total measurement uncertainty.

2.1.3.4 Charge Question 1.3d. *Please comment on the concerns of stakeholders that calculating measurement uncertainty for field measurements makes the survey process difficult to implement.*

Quantifying measurement uncertainty for field measurements (which is understood to be for scan-only and in situ measurements), is important, most especially for documentation of final status survey (FSS) results. The scan-only approach in which a large fraction of the site to be surveyed is measured in theory captures mean concentration much better than would be achievable with sampling, so long as the measurement characteristics of the scanner (and operator) meet appropriate criteria. The factors affecting the performance of scan-only designs are largely detailed in Chapter 6 and would seem to be useful to the stakeholders. Also, after implementing the recommendations from the panel (with more detailed explanations and worked examples) the revised MARRSIM revision 2 should be clearer and easier to use in calculating measurement uncertainty. It is noted on page 6-27 of the draft MARSSIM document that rigorous uncertainty assessment for field measurements is generally only necessary for final site survey documentation, and generally not required in scoping or characterization surveys. Surveys prior to the FSS will need to exhibit good detection ability, but issues regarding calibration uncertainty can be postponed to the FSS.

2.1.3.5 Charge Question 1.3e. *Please comment on whether recommendations provided by NIST, ANSI/IEEE and MARLAP for measurement quantifiability should be incorporated further into MARSSIM, Revision 2, or whether the current recommendations should be left as is (e.g., the original MARSSIM requirement that the MDC/MDA should be set at 10-50% of the action level).*

The discussions in MARLAP, and other MAR series manuals, around MDC/MDA, focus on Poisson variability in individual laboratory measurements (counts) without consideration of the variability of the concentrations over the site being investigated. The latter source of variability was seen by MARSSIM Revision 1 (U.S. EPA, 2000) as the dominant source of uncertainty affecting study design and relatively little was said about the calculation of MDCs. Poisson variability for scan-only instruments used over the entire survey site would likely shrink to negligibility because the average of a very large number of readings taken by the scanner will exhibit very little such variability. However, the SAB recognizes there may still be requirements to validate the scan-based methods against radiochemical laboratory-based readings for which the additional Poisson counting material in MARLAP is at least partly relevant.

Regarding the issue of controlling for quantifiability versus detection, the expanded uncertainty with a stated coverage factor when it is based on factors relating to the DQOs controls the distribution of possible values around the MDC. The MQC is a different method of controlling the uncertainty of measured values usually at the UBGR and below. Either of these concepts (MDC or MQC) can be used for determining an MQO of optimal detection, but not both simultaneously;

either method will provide a satisfactory control on the distribution of values measured in the gray region as long as consideration is given for the DQOs and the methods of analysis.

Regarding the issue of the setting of quantifiability criterion, the SAB agrees with keeping the original MARSSIM requirement of a measurement method with MDC equal to between 10 and 50% of the UBGR so long as they relate to the data quality objectives, e.g., the tolerable error rates and the width of the gray region. (Section 4.8.2 of the draft MARSSIM document is less explicit saying only that the MDC should be less than 50 percent of the UBGR region. This should have little practical effect on study design compared to the original MARSSIM requirement.) It is noted in Appendix C of MARLAP that setting the MDC to 10 to 50 percent at the action level is tantamount to setting the relative standard deviation to between 0.03 to 0.17, i.e., roughly 10 percent at the action level, assuming a tolerable error rate of 5 percent for Type I and Type II errors. Therefore, in this case the MQC will be close to the action level. Since the two criteria (setting the MDC to between 10 to 50 percent of the action level, vs. setting the MQC to give a relative standard deviation of 0.10 at the action level) are nearly equivalent, it seems that either criterion will function well as an MQO. Of course, there may be situations when this requirement is not strict enough and either the required MQC or the MDC should be strengthened (i.e., lowered). The SAB agrees with MARLAP (Chapter 3 page 3-12 to 3-13) that the value of including the MQC as a possible performance characteristic is to emphasize the importance of the quantification capability of a method for those instances where the issue is not whether an analyte is present or not, but rather how precisely the analyte can be quantified. The case of quantifying low levels of ^{238}U in soil is mentioned, for which presence of the analyte is expected in the background. This discussion could be expanded (perhaps into a full example) if a realistic scenario can be found.

Overall Recommendations for Charge Question 1.3:

- The subsection on the first MQO, Measurement Method Uncertainty in Section 4.8.4.1, should be expanded, with clear descriptions of the concepts of Measurement Method Uncertainty, Required Method Uncertainty, the calculation of the Required Method Uncertainty, and its use in the DQO process.

- Include detailed, step-by-step, worked out examples of setting up σ and u_{MR} (this is the required method uncertainty as defined in MARLAP) uncertainty component lists for a few cleanup scenarios with the expected measurement systems to be used. This would make this document easier to implement for real projects.

- Specific, detailed examples of how to calculate the CSU for different measurements/methods are needed.

- A fully worked out example of determining an MDC should be provided in Chapter 6.

- It should be emphasized that regardless of the difficulty of calculating or estimating uncertainty of a measurement, each measurement should have an associated uncertainty that is used in the final standard combined uncertainty.

- Use of standard statistical terms, such as Type A and Type B uncertainties, and specific examples of how they are used in the case studies will provide excellent guidance for the user. This also includes use of coverage factors, sensitivity coefficients, correlated/uncorrelated uncertainties.

- A measurement equation model for the selected survey scenarios and measurement processes should be included as part of the examples.

- The manual should give due consideration regarding the overall definition of MDC or MDA because the manual references many measurement methods that do not follow the normal counting statistics used in the rudimentary examples given in Chapter 6.

- The descriptions in MARSSIM of measurement uncertainty, MDC, MDA and detection should be the same as they are in MARLAP as the two documents are complementary. Exceptions are permitted in MARSSIM to accommodate field measurement techniques that differ from laboratory analyses of collected samples but should be accompanied by a clear description of the rationale.

- If the uncertainty process is too difficult to implement in some *very limited circumstances*, a summary table of various approaches or considerations that might be employed instead could be included. Such a table would assist in data analysis and decision making and may be a useful addition.

- In order to ensure that both non-correlated and correlated uncertainties are appropriately accounted for, the description for Equation 6-18 should reference Equation 19-11 from MARLAP.

- MARSSIM should add to Section 6.2.2.3 that spike samples and standards need to be confirmed for commutability, i.e., that they are representative of the areas being surveyed.

- A small section should be added on how to perform the assessment of final measurements and their uncertainties with regards to the MQOs, and what actions need to be taken if the MQOs are not achieved.

- The original MARSSIM requirement of a measurement method with MDC equal to between 10 and 50% of the UBGR should be used in setting the quantifiability criterion.

2.1.4. Charge Question 1.4. Unity Rule

2.1.4.1 Charge Question 1.4a. *Is the discussion of survey requirements for areas of elevated activity technically accurate, appropriate and clear?*

The SAB finds the discussion of survey requirements for multiple areas of elevated areas of radioactivity to be overall technically appropriate and useful, but ambiguous in some instances. For large survey areas, MARSSIM assumes a relatively uniform distribution of radioactive material.

The average concentration of radionuclide of concern over the entire area is used to determine if the concentration is below the Derived Concentration Guideline Levels ($DCGL_w$) using statistical tests (see response to Charge Question 1.2). For small areas of elevated concentrations of radioactive material, MARSSIM recommends simple comparison of the results of individual measurements to an investigation level, $DCGL_{EMC}$, using the elevated measurement comparison (EMC) approach. Individual concentration is compared to the $DCGL_{EMC}$ (the Derived Concentration Guideline Level used for EMC) for compliance evaluation. The $DCGL_{EMC}$ is derived (by the licensee or operator) separately for these small areas, generally using different exposure assumptions than those used for large areas and is $\geq DCGL_w$. Any measurement from the survey unit that is equal to or greater than the $DCGL_{EMC}$ indicates an area of relatively high concentrations that should be investigated. If elevated levels of residual radioactive material are found in an isolated area in addition to residual radioactive material distributed relatively uniformly across the survey unit, the Unity Rule (discussed below) can be used to ensure that the total dose or risk meets the release criteria. If there is more than one of these areas, a separate term should be included in the calculation for each area of elevated activity. The SAB finds the EMC approach appropriate and accurate. The $DCGL_{EMC}$ can be higher than the $DCGL_W$ due to the lower dose or risk resulting from exposure in a smaller area of radioactive material. This concept of $DCGL_{EMC}$, highlighted elsewhere in the draft revision, needs clarification in Section 8.6.1 to enhance the document. The survey requirements in Chapters 4 and 5 provide a practical and comprehensive approach to completing a radiological survey of surface soils and building surfaces. Additional details would be helpful, including more detailed examples. Examples 7 and 8 in Section 5.3.5 on pages 5-39 and 5-40 provide the calculation for the number of chosen samples. Greater understanding could be achieved if additional details, such as showing all the steps, including assumptions to simulate a real case, were provided and the context broadened.

Clarity in the survey development process (page 5-36, line 30), can be achieved if the narrative provides insight into the decision on selection of a triangular or rectangular grid for systematic sampling. There are certainly advantages/disadvantages to both. These should be briefly presented before referring the reader to another EPA publication (U.S. EPA, 1994) for an in-depth explanation.

2.1.4.2. Charge Question 1.4b. *In particular, please comment on the decision to maintain the use of the Unity Rule for multiple areas of elevated activity (Section 5.3.5, Section 8.6 and Appendix O.4).*

The Unity Rule is used to ensure that the total dose (risk) from all sources (or media) and all radionuclides associated with each source does not exceed the release criteria. It is to be used when more than one radionuclide is present and distinguishable from background and a single concentration does not apply. Essentially, this means that if measurements of different quantities are made at a location, then the Unity Rule must be used. For example, the Unity Rule would be used if two radionuclides are measured in each soil sample or if gross alpha and gross beta measurements are made at each location and the results are being compared to specific DCGLs.

$$\sum_{i=1}^{n}\left(\frac{\text{Dose or Risk Component}}{\text{Release Criterion}}\right)_i = \sum_{i=1}^{n}\frac{C_i}{DCGL_i} \leq 1 \quad (4\text{-}5)$$

where

- C_i is the concentration of the ith component (e.g., radionuclide or source) leading to dose or risk.
- $DCGL_i$ is the derived concentration guideline level of the ith component (e.g., radionuclide or source) leading to dose or risk.

If residual radioactive material is found in an isolated area of elevated activity—in addition to residual radioactive material distributed relatively uniformly across the survey unit—the unity rule (above) can be used to ensure that the total dose is within the release criteria, as shown in Equation 8-4:

$$\frac{\delta}{DCGL_W} + \frac{(\text{mean concentration in elevated area} - \delta)}{DCGL_{EMC}} \leq 1 \quad (0\text{-}2)$$

where

- δ is the estimate of the mean concentration of residual radioactive material in the survey unit.

If there is more than one elevated area, a separate term could be included in Equation 8-4 for each area. The use of the Unity Rule for more than one elevated area implies that a person is exposed simultaneously from each elevated area.

The SAB believes the Unity Rule should be maintained in MARSSIM Revision 2. The rule provides a consistent and conservative approach for a reasonable dose assessment.

While there are significant limitations to the Unity Rule (see Section 8.6.2), the SAB believes the use of the Unity Rule, will yield a conservative (protective) result. The Unity Rule limitations include:

- The implication that a person is centered on each area of elevated radioactive material and exposed simultaneously when there are more than one elevated areas.
- This simultaneous exposure at many elevated areas will overestimate doses/risk level.

In addition to the limitation noted above, the use of the Unity Rule when discrete radioactive particles (DRPs) are present is not recommended. DRPs are point sources, usually on the order of millimeters to micrometers in size. They often consist of more than one radionuclide and because of the intense electromagnetic field generated by the high activity density, the emissions of all the radionuclides can be anisotropic. This characteristic can make the DRPs easily missed when performing scanning surveys. DRPs also dry out fast and take on an electrostatic charge, allowing them to be very mobile and thus overlooked. When the historical site assessment (HAS) identifies the possibility of DRPs and one is found, that is a strong indication that there will be others. Removal of DRPs should occur soon after their discovery during characterization surveys. DRPs do not conform to the concept of dispersed area or volumetric sources assumed by the risk pathway models included in either the $DCGL_W$ or $DCGL_{EMC}$ situations. Since DRPs are classically treated as point sources, the guidance established for area sources or volumetric sources used in MARSSIM is not appropriate for DRPs. Treating the area around a DRP as an EMC creates an extreme sampling situation due to the small area created by the microscopic DRPs. The Unity Rule is used for dose assessment (and compliance with dose limits), averaging out area doses based on occupancy over the entire area, and fails to take into account the complexities of DRPs.

As an initial default position, the Unity Rule is preferable to another (weaker) rule that may be easier to use but is not conservative and leads to inappropriate site release. In situations where multiple radionuclides exist, in an unrelated or non-equilibrium state, the Unity Rule would provide a more reasonable estimate of dose since each would have a different $DCGL_{EMC}$. The use of the Unity Rule is consistent in these uses with other agency rules. A fuller derivation of the Unity Rule in Section 4.4 (or in an appendix) is recommended. Additionally, examples of the EMC Unity Rule should be provided: one showing acceptance (<1) and one showing failure (>1).

Coordination with the applicable federal or state regulators is encouraged. Using the Unity Rule to ensure total dose is within the release criteria can cause unnecessary effort and expense. As shown in Abelquist (2008), doses from elevated areas other than a single primary area can be relatively small and negligible. Additional consideration from a second (or additional) elevated area may require changes to the underlying exposure scenario (e.g., percentage time at a given location). To accomplish these changes, the regulatory agency will need to be notified, briefed and concurrence obtained. Section 4.5.2 indicates the implementer can justify to the regulator when no $DCGL_{EMC}$ requirement is needed. In most cases, early remediation of multiple elevated areas will circumvent the need to use a complex unity formula with additional terms addressing individual elevated areas.

In the narrative in Chapter 8 Section 6, use of the Unity Rule is suggested to ensure the total dose is within the release criteria. Using the Unity Rule for a single elevated primary area is a good decision, however, the Unity Rule for two or more elevated areas has a potential to cause an overly conservative (and potentially impossible) scenario. These types of scenarios lend themselves to alternative approaches, which are alluded to in the text. The user needs to be well versed in the modeling software being used to develop DCGL's. When defaults are no longer properly used (e.g., resident time as a percentage at a location), the user needs to discuss the issue and obtain concurrence from the regulatory body overseeing the remediation and release of the building and/or grounds.

2.1.4.3. Charge Question 1.4c. *Are there suggested alternatives to the use of the Unity Rule?*

A critical criterion for any alternative to the Unity rule is whether it can *accurately reflect* the potential dose and thus compliance for site release. Some panel members conducted a limited literature search on this question but had no success in finding a well-documented alternative.

The only alternatives cited were those provided in Section 8.6.2 of the draft MARSSIM document:

1. The MARSSIM user could determine the elevated area (primary area) that contributes the most to the total dose or risk. As shown by Abelquist (2008), the doses from elevated areas other than the primary area can be very small and might be negligible.
2. The dose or risk due to the actual residual radioactive material distribution could be calculated if an appropriate exposure pathway model is available.

Comparing the two alternatives provided by the draft MARSSIM document, the first alternative relies on an assumption that the doses from other than a primary area are relatively insignificant. If the assumption is true, this implies the Unity Rule would not be constrained by simultaneous two (or more) location exposure concerns and thus reducing unnecessary remediation efforts. The second alternative would always be an option, providing that it can be implemented using sufficient characterization data about the pathways of interest.

Recommendations:

- The concept of $DCGL_{EMC}$, needs clarification in Section 8.6.1 to enhance the document. Additional details would be helpful, including more detailed examples.

- Greater understanding of the $DCGL_{EMC}$ could be achieved if additional details, such as showing all the steps, including assumptions to simulate a real case, were provided and the context broadened.

- Clarity in the survey development process (page 5-36, line 30), can be achieved if the narrative provides insight into the decision on selection of a triangular or rectangular grid for systematic sampling.

- DRPs should be removed before applying the Unity Rule to a site.

- The SAB believes the Unity Rule should be maintained in MARSSIM Revision 2. The rule provides a consistent and conservative approach for a reasonable dose assessment.

- Detailed examples of using the EMC Unity Rule should be provided: one showing acceptance (<1) and one showing failure (>1).

- Coordination with the applicable federal or state regulators is encouraged. Without regulatory guidance, using the Unity Rule to ensure total dose is within the release criteria can cause unnecessary effort and expense.

- In lieu of the Unity Rule, if actual radioactive material characterization data are available, site-specific exposure dose/risk calculations can be performed using an acceptable exposure pathway model.

2.1.5. Charge Question 1.5. Discrete Radioactive Particles

Is the discussion of the use of MARSSIM surveys for addressing sites containing discrete radioactive particles technically sound and appropriate, and is the description accurate? In particular, please comment on the rule-of-thumb for determining when use of MARSSIM may not be appropriate for survey units containing discrete radioactive particles (Section 4.12.8 and Appendix O.5).

The SAB finds the discussion of DRPs currently provided in the draft MARSSIM document to be inadequate and should be augmented with information described below**.** DRPs, sometimes called hot particles, specks or fleas, refer to very small (usually on the order of millimeters to micrometers), highly radioactive particles. The draft MARSSIM document devotes one section only to the subject of DRPs (Section 4.12.8) that is two paragraphs in length. This section is reproduced verbatim in Appendix O.5. Therefore, the Appendix provides little added value. The draft MARSSIM limits its discussion of DRPs to an appropriate and useful cautionary statement advising against using the Elevated Measurement Comparison (EMC) process when DRPs are discovered. However, the presence of DRPs creates issues not discussed in MARSSIM that expand beyond the avoidance of the EMC survey design and concomitant data analysis processes. The absence of information regarding the impacts DRPs will impart to the overall site survey process makes MARSSIM incomplete and less useful for controlling the special risks when such sources are present. DRPs pose unique hazards with implications that require more visibility and consideration within the Manual. Presently, the draft MARSSIM document focuses on avoiding impractical survey designs through the use of an inadequate rule of thumb aimed at distinguishing a point radioactive source from a source of radiation arising from radionuclides distributed across a large area such as on the surfaces of buildings or within a large volume of surface soil. Elevated areas of radiation trigger special survey designs to assess whether the elevated concentrations of radioactive material exceed Derived Concentration Guideline Levels for Elevated Measurement Comparisons ($DCGL_{EMC}$).

Dose pathways are different for DRPs compared to volumetric or bulk radioactive contamination of building surfaces and surface soils. The concept of DCGLs as used in MARSSIM is unlikely to be effective for DRPs. DRPs present varied risks depending on the radionuclide composition, activity level and prevalence at the site. DRPs tend not to be singular and the discovery of multiple DRPs should be anticipated. General guidance has not been sufficiently developed and the treatment of DRPs will likely entail site-specific negotiations among the stakeholders as to the risks posed by DRPs, the level of effort to discover poorly detectable particles and the consequences of leaving DRPs in place.

Some DRPs may present significant health hazards. The high radioactivity levels and associated dose rates from some radionuclides known to comprise DRPs may, depending on the duration of exposure, induce tissue reactions (formerly known as non-stochastic or deterministic effects) should a DRP become attached to the skin or be introduced to the internal organs of the body. More information regarding tissue reactions can be found in ICRP Publication 118 (ICRP, 2012). The National Council on Radiation Protection and Measurements recommended limits for skin exposure to "hot particles" in 1989 (NCRP, 1989) following the Nuclear Regulatory Commission's issuance of Information Notice 87-39 (NRC, 1987) about controlling hot particles at nuclear power plants. The Nuclear Regulatory Commission has made available a computer code for calculating the skin dose arising from various contamination scenarios as an aid for demonstrating compliance with regulatory dose limits (NRC, 2018).

The presence of DRPs will impact many planning decisions, quality assurance considerations and survey procedures. The ability to detect DRPs depends on the radionuclide composition, the radioactive emission rate, the radiation types emitted, the shielding effects of soil thickness and surface coatings, and measurement instrument capability. Scanning methodologies will most likely be the preferred means of discovering DRPs causing operator performance to be a consideration. DRPs will influence the process and selection of DQOs and Measurement Quality Objectives (MQOs) as well as requiring appropriate radiation protection procedures. A site contaminated with DRPs cannot conform to the basic tenets contained in the draft MARSSIM document. DRPs do not distribute uniformly and the radionuclides contained within a DRP cannot be relied upon to exhibit the same type and ratios of radionuclides expected to be more widely dispersed across the site.

DRPs can be extremely mobile presenting possible contamination concerns. The high radioactivity levels associated with some DRPs can impart an electrostatic charge causing DRPs to be very mobile potentially causing them to jump to adjacent areas, attach to clothing, and contaminate instruments thus leading to the term, fleas. DRPs may be difficult to detect depending on the composition and quantity of radionuclides and their emitted radiation qualities. This will influence instrument selection and the development of investigation or action levels. It will be important to understand instrumentation characteristics to enable the discovery and localization of DRPs on personnel and equipment as such discovery during the operational radiation protection program may be useful indicators of risk.

The rule of thumb presented as Equation 4-26 in the Manual may be difficult to implement and may not have the intended effect of avoiding the use of the EMC process. The draft MARSSIM document introduces a rule of thumb to identify those conditions in which the discovery of a DRP would preclude the use of the EMC process. The rule features equation 4-26 from Section 4.12.8

$$d > 3L \qquad (4\text{-}26)$$

where L is the estimated longest dimension of the area of elevated activity, and d is the distance to the detector.

The rule requires the distance, *d*, between the DRP and detector to exceed by three times the longest dimension of the area of elevated activity, *L*, created by the DRP. This relationship between *d* and *L* attempts to describe the condition when a source would appear to a detector as a point source; however, there are several practical impediments to using a rule of thumb of this type. First, given the size and unknown location of the particle, it appears that *d,* the distance between the source and detector, would be difficult to establish; particularly with instruments that present several square centimeters of active detection area. The DRP may reside at some unknown depth in soil or in a crack in a building surface. Secondly, L must be established according to some quantitative criterion that clearly defines an area of elevated activity. DRPs may exist in a larger area of elevated activity created by the volumetric or areal dispersion of other radioactive material. The area of elevated activity around a DRP will depend on the types of radiations emitted; radionuclides emitting gamma rays are able to create an elevated area much larger than those emitting beta particles. Many DRPs consist of a mix of radionuclides emitting both gamma and beta particles to create a complex pattern of measured elevated activity. DRPs consisting of radionuclides emitting primarily alpha particles will be extremely difficult to discover such that neither *d* nor *L* can be ascertained reliably if at all in many field conditions. One can postulate conditions for *d* and *L* for a DRP that would fail to avoid the use of the EMC process. Thirdly, the rule assumes that a single particle may be responsible for the elevated reading when in fact multiple DRPs may be present.

Recommendations:

- The unique issues created by DRPs should be introduced early in the MARSSIM document, preferably in Chapter 2 that introduces the main concepts and the overall order of information contained in the document, but certainly by Chapter 3 that deals with the Historical Site Assessment (HSA) during which the potential existence of DRPs will most likely be discovered. Sections of subsequent chapters should address specific actions or decisions impacted by the need to consider DRPs.

- Decision aids, perhaps site-specific guidelines, would be beneficial when additional treatment of DRPs is warranted. Such decision aids will allow resources to be directed to those aspects most important in terms of dose and risk. Dose and risk criteria are beyond the scope of the MARSSIM document, but some means are needed on which to judge the priority required for addressing the presence of DRPs.

- The MARSSIM document should call special attention to the implementation of radiation protection controls to prevent excessive doses to workers or others present during surveys and remedial activities at sites expected to contain DRPs. At a minimum, the hazards potentially presented by DRPs should be discussed in Section 4.10. Operational insights gained from records of radiation protection programs enacted during facility operation and/or decommissioning may suggest the likelihood of discovering DRPs.

- Given the potential health hazards arising from contact exposures to DRPs, survey activities aimed at discovering particles deemed to be hazardous should be conducted early in the sequence of survey types, for example, as part of the scoping surveys of areas likely to contain DRPs. According to Section 2.2.2 of the MARSSIM document, areas of sites expected to contain DRPs will be appropriately identified as Class 1 invoking the necessary planning and 100% survey coverage. If discovered, DRPs should be removed promptly and the remediated area rescanned to validate removal. Removal of DRPs may enable the EMC process to be appropriately used without the presence of DRP measurements affecting the statistical data analyses that could result in the inadvisable use of the EMC process. Prompt removal may also avoid the need to develop and use site-specific rules of thumb as to whether the EMC process is advisable.

- The discussion of the DQO and MQO processes in Chapter 4 should address the influence DRPs will exert on the requirements for measurement methods and their uncertainties. Detection and uncertainty requirements will affect instrument selection and survey procedures. DRPs are most likely to be discovered by scanning techniques and the DQO and MQO processes should identify instrument response patterns that would indicate the presence of DRPs. A figure devoted to summarizing the planning processes for DRPs would be beneficial. If historical analyses or preliminary surveys reveal the presence of DRPs, then the Measurement Quality Objectives for all remaining survey activities should anticipate the measurement of area/volumetric and point sources. This naturally leads to establishing separate measurement detection limits and measurement uncertainty requirements that may have an effect on the selection of instrumentation and scanning procedures.

- The MARSSIM document should include cautionary statements about the mobility and contamination risks posed by DRPs.

- An alternate rule of thumb or rules of thumb should be evaluated. The aim of the rule is to identify simple measurement data available from scans that indicate the presence of small, discrete radioactive particles that either individually or collectively pose a material risk if not identified and removed. One possible approach might be to establish a threshold distance between the point or location a detector reads a maximum count rate and that point when the detector reads a fraction of that maximum, perhaps one-half the maximum. If this change in count rate occurs over a small distance as might be expected from a small source, then a DRP might be present and a targeted investigation initiated.

2.2. Charge Question 2. Technical Approaches and Examples

Does MARSSIM, Revision 2 provide useful, appropriate and clear examples and descriptions of technical approaches to implementing surveys and the statistics by which they are interpreted?

The SAB finds technical approaches and statistics for implementing surveys are generally useful and accurate, but recommends improvements in Chapter 6 to:
- incorporate uncertainties in the C conversion factor of the Minimum Detectable Concentration (MDC),
- acknowledge when differences between ideal and realistic conditions merit specific treatment in the technical approaches,
- provide updated statistical techniques for modern data-logging systems that no longer rely on human surveyor data interpretation to perform the survey, and
- include updated information on measuring radon and its decay progeny.

Additionally, the SAB finds that Appendix H is useful and generally accurate in communicating valuable summary information but could be improved by a more consistent presentation. Several suggestions are provided to this end.

Regarding the examples in Chapter 6, the SAB recommends that Example 9 be improved with a more complete representation of combined uncertainty that includes uncertainties in instrument efficiency. In Section 6.4.4 on confidence intervals, Example 8 on uncertainty propagation is so simple that it is unlikely to be representative of a real project for which many different sources of uncertainty must be determined. Inclusion of a more complete example would be much more instructive by taking the reader through the whole process of calculating measurement uncertainty in implementing site surveys.

Regarding the use of Ranked Set Sampling for hard-to-detect radionuclides, the idea is to use some alternative "easy-to-measure attribute" which is highly correlated with the substance of interest to rank the sites by their projected level of contamination, and thereby to select a subset of sites to measure the hard-to-detect substance, in the hope that this will give a more representative distribution of measurements than would be achieved by simple random sampling (SRS). However, Appendix E has not included discussion on how highly correlated the attributes need to be, and how should one assess, in practice, whether the conditions for RSS are met.

The SAB finds two examples in Appendix E not useful, appropriate, and clear. Example 1 shows that of the 12 sites for which both laboratory analyses and field screening measurements are given, the sample correlation coefficient of the two sets of measurements is 0.998. This leads to the suspicion that this is just a made-up example, not even based on any real dataset. If this is true, that might still be acceptable for the purposes of illustrating the sampling method, but the report should acknowledge that fact and make clear that real data sets are extremely unlikely to display such high correlations.

Example 5 is a much more carefully laid out example with a lot of specific experimental detail, but it still raised concerns where the example came from. While clearly working from a real dataset, the concern is whether or not Example 5 is a one off, or whether many such datasets (for different analytes) could be developed. The specific concern is whether many pairs of easy-to-measure and hard-to-measure pairs with a high correlation between the two can be adduced in this setting. If only a few realistic examples could be developed then this raises a question about the relevance of the approach to cleanup and FSS determination, and in particular whether a lengthy discussion of the method is worthwhile. While the Ranked Set Sampling method is widely known in some fields (e.g., Ecology), the one realistic example provided in the draft MARSSIM Revision 2 did not increase the confidence of the SAB that this method is helpful enough to be included in the MARSSIM revision 2.

The SAB finds the examples in Chapter 5 of the draft MARSSIM document to be useful, appropriate, and clear. However, many of the examples should include additional detail on the calculations and use of tables, and more references to other appropriate sections of the MARSSIM document. About half of the examples are unchanged from MARSSIM, Revision 1(U.S. EPA, 2000) and half are modified. There is one new example that was needed to support the process for implementing Scenario B. Overall, the reformatting and modifications of examples in Chapter 5 are a marked improvement over examples in MARSSIM, Revision 1, but additional detail would be useful for better clarity.

2.2.1. Charge Question 2.1. Measurement Methods and Instrumentation

Please comment on whether the description of updated measurement methods and instrumentation information (Chapter 6 and Appendix H) is useful, appropriate and clear.

Chapter 6:

The description of measurement methods and instrumentation information in Chapter 6 is generally useful and in large part appropriate and clear. The content of Chapter 6 is laid out well with the inclusion of excellent tables, examples and sample calculations. The chapter presents common sense approaches for how to choose the best detection method for a given radionuclide that may emit more than one type of radiation as well as an excellent discussion of measurement uncertainties and how to manage them. However, there are descriptions of concepts and operations that could be made clearer and more useful.

The SAB finds that the treatment of uncertainties in various quantities needs improvement for clarity and completeness. First, it needs to be made clear to the reader that the minimum detectable concentration (MDC) is an *a priori* estimate. NUREG-1507 Revision 1 [Nuclear Regulatory Commission (NRC), 2020] clearly shows an MDC calculated under ideal conditions should be considered as providing information on the general detection capability of the measurement system and not as absolute levels of activity that can or cannot be detected [in the field]. Equation 6-5 only applies to detection systems that operate in pulse or counting mode; it is not applicable to current mode detection systems, such as an ion chamber or thermoluminescent dosimeter (TLD). Furthermore, equation 6-5 is only applicable when the sample is to be counted for the same length of time as the blank.

It is properly stated (p. 6-9, lines 14-26) that the MDC is dependent on "C," the factor that converts the Detection Limit from blank/background counting signal to concentration measurement. This factor is composed of several subfactors that have associated uncertainties (e.g., surface type, source efficiency, source-to-detector geometry, source count time, and mean counting efficiency).

It is recommended that Equation 6-5 be rewritten to include the uncertainty in C and provide an example illustrating how neglecting these uncertainties can lead to an erroneous MDC. Clarification is needed to differentiate MDC in units of Bq/kg and MDC in units of Bq/m^2. Surface (areal) concentration would be reported in units of Bq/m^2 while soil concentration would be reported in units of Bq/kg. Designation of concentration units of Bq/m^2 and Bq/kg will inevitably lead to confusion. The SAB recommends Chapter 6 address surveying soil that is volumetrically contaminated with a radionuclide that emits medium energy gamma-rays to highlight differences in the concentration (Bq/kg) compared to areal concentration (Bq/m^2).

It is stated (P. 6-17, lines 5 - 6) that the surveyor efficiency was estimated between 0.5 and 0.75, and the draft MARSSIM recommends a value toward the lower end of this range (i.e., 0.5) for estimating the Scan MDC. The SAB recommends explaining how an implementer should determine the "surveyor efficiency" and how to assign uncertainty. Scanning surveys of land areas are addressed within Section 6.3.2, and the minimum detectable count rate ($MDCR_{surveyor}$) is infused in a majority of the equations and examples. However, it should be clarified that $MDCR_{surveyor}$ is not necessary for data-logging detection systems. Because data logging is included and available in modern detection systems, current guidance pertaining to land area surveys in the draft MARSSIM document is outdated. Many of the fundamental statistical principles introduced in Section 6.3.1 still apply, so a minor revision of this section is highly recommended. The statement on scanning techniques in Section 6.6.1.1 reinforces this point, "Scanning equipment coupled with GPS or other locational data is strongly recommended for scan-only surveys."

Regarding recommendations in ISO-7503-1 (ISO, 1988), source efficiencies (page 6-40, lines 35-40) should include uncertainty estimates, and nonstatistical uncertainties [also referred to as Type B by the National Institute of Standards and Technology (NIST, 1994)] should be incorporated into the determination of standard combined uncertainties. Monte Carlo simulations that take into account source related factors such as type of radiation and its energy, source uniformity, surface roughness and coverings, and surface composition (e.g., wood, metal, concrete) could be helpful in estimating the source efficiency. Regardless, estimating a source efficiency without actual measurements should be accompanied with an uncertainty estimate that is to be incorporated into the standard combined uncertainty of the measurements made by the measurement systems. At a minimum a rectangular uncertainty distribution could be used.

Example 9 (Page 6-51) provides an opportunity to demystify handling the concept of uncertainties. The SAB recommends including how the estimated uncertainty of the overall efficiency would be calculated and mentioning the type of instrument corresponding to the stated instrument efficiencies. Correcting for process blank, decay to a reference time, and radiation emission probability in the conversion of counts to the derived concentration guideline level (DCGL) in soil should be included in the following section (page 6-52, lines 3-5).

Uranium and thorium in the soil as a source of radon should be included in the discussion. The effects of radon and its progeny on fixed and survey measurements of other radionuclides should be discussed as a source of measurement interference and mentioned with the potential for contamination of scanning instrumentation. The discussion of estimating contributions from radon exposure should be part of the determination of the action level (AL) and DCGL.

More should be said about the needs to recalibrate when field conditions change (page 6-38, lines 32-33). Calibrations should consider differences in field conditions and influences from environmental changes during the survey. Either a recalibration should be performed for each measurement system using commutable certified reference sources that accommodate the new measurement conditions, or when possible, computational bias corrections and uncertainties estimates should be applied to the measurement systems to accommodate the new effects. Furthermore, calibrations should be checked at a prescribed interval including use of quality control check sources to assure continued and stable operability between calibrations and compliance with the MQO requirements.

There is a good discussion recognizing the many factors that complicate radon measurements. Progress could be made by recognizing these complicating factors in the estimation of measurement uncertainty. The reference to Jenkins 1986 (Page 6-53, line 14) could lead readers to believe that there have been no new developments over the last 3 decades for measuring radon and radon progeny in air. The SAB compiled an updated list of references at the end of this section. These updated reference(s) should be included in the MARSSIM Revision 2 manual.

Additional comments for improvement of Chapter 6 are listed below:

- Page 6-26, Section 6.4 Measurement Uncertainty – Comments regarding vocabulary associated with uncertainty made in response to Charge Question 1.3 should be implemented in this section.

- Page 6-35, lines 36 – Cadmium telluride is not a scintillator and should be removed from the list or replaced with the scintillator that the MARSSIM authors were intending, e.g. $CdWO_4$.

- Page 6-36, line 9 – Reference is made to cadmium zinc telluride (CZT), but cadmium telluride (CT) is commercially available. Recommend adding CT to the discussion.

- Page 6-37, line 2 – Recommend adding a sentence on the sensitivity of TLDs.

- Page 6-37, line 11 – Recommend adding a sentence on the capability of the electret ion chambers (EIC).

- Page 6-41, line 14 – A more appropriate background exposure rate would be 10 μR/h.

- Page 6-46, Table 6.7 – Usage of "Good", "Fair", and "Poor" seems subjective. There should be text that explains how these ratings were determined. Also, the basis for downgrading of the instruments for the scanning surveys should be added.
    - Table 6.8 provides useful information, but its presentation could be more effective if divided into two tables. There are advantages and disadvantages to instruments, and there are advantages and disadvantages to measurement technique; it becomes very repetitive to try to address both in a single table.
    - "Hand-Held Instruments" "Direct" "Advantages", 3rd bullet – refers to the ability to efficiently measure alpha, which is inconsistent with Table 6.7.
    - "Hand-Held Instruments" "Scanning" "Advantages", 3rd bullet – refers to efficiently measure neutron radiation, which is inconsistent with Table 6.7.

- Page 6-56, Table 6.9, 1st row – The SAB is not aware of putting activated charcoal into liquid scintillation cocktail; the color quench must be pretty significant. "Not a true integrating device" needs clarification. This is also mentioned on Page 6-58 without explanation.
- Page 6-59, Line 4 – As mentioned in the previous paragraph, electrostatic attraction of radon progeny to the surface of a detector is also an option.

Recommendations for Chapter 6:

- Incorporate uncertainties in the C conversion factor of the MDC.

- Acknowledge when differences between ideal and realistic conditions merit specific treatment in the technical approach.

- Provide updated statistical techniques for modern data-logging systems that no longer rely on human surveyor data interpretation to perform the survey.

- Include updated information on measuring radon and its decay progeny.

- Revise Example 9 with a more complete representation of combined uncertainty that includes uncertainties in instrument efficiency.

- Recommend adding more discussion on available optically stimulated luminescence (OSL) materials on page 6-37, similar to what is done with the TLD materials, and including a sentence on the sensitivity of OSLs.

Suggestions for Chapter 6:

- Page 6-31, lines 2 - 3 – Suggest including radioanalytical service representatives in addition to service providers to perform field data collection activities. MARLAP (U.S. EPA, 2004, Chapter 5 on Lab Services) provides guidance for selecting a service provider, including pre-award proficiency evaluation, assessment and audit. An appropriate accreditation program, based on MARSSIM Revision 2, could provide interagency consistency among capabilities, operations, quality of results and relieve programs from conducting these quality improvement tasks on a case-by-case basis.

- Page 6-34, lines 25 - 28 – The metrologists and subject-matter experts can address situations other than radionuclides uniformly distributed on a plane or through a regularly shaped volume (e.g., a disk or cylinder). Reference MARLAP (U.S. EPA, 2004) and mention that traceable secondary reference laboratories and source producers should establish DQOs and MQOs for appropriate calibration sources.

- Page 6-39, lines 14-24 – MQOs need to be specified with the Calibration Source Provider to assure the appropriately commutable certified reference standards are prepared for accurate calibrations of direct measurement systems within necessary combined standard uncertainty limits. Commutability relates to how well reference materials represent actual conditions during the survey in addition to providing certified values and uncertainties for analytes in certified reference materials.

Appendix H:

Appendix H could be improved by a more consistent presentation of pros/cons/applications in the text and tables. Sometimes the information is in the Description column, sometimes in the Applications column and sometimes in the Remarks column. Additional effort on a consistent organization would be beneficial to readers. Furthermore, consider technical editing of Appendix H. For the most part, the descriptions of the individual instruments are good, but there are inconsistencies.

To the extent practicable, listing prices in 2020 dollars is suggested. Some of the prices seem low, and others are identical to the previous version of MARSSIM, which may no longer be accurate. Although discussion on unavailable technologies in Appendix H is not immediately useful, it is appropriate and should remain because technological advancements and market forces over time tend to transform some unavailable technologies into widely available technologies. Specific suggestions for improving Appendix H follow.

Specific comments on Appendix H:

- Page H-1, line 26 – There can also be market contraction.
- Page H-1, lines 14 and 79 other occurrences in Appendix H – Appendix H refers to "sensitivity" while Chapter 6 refers to "capability". Appendix H should refer to "capability" so that there is consistency within MARSSIM and consistency with MARLAP (EPA 2004).
- Page H-10 – The difference between this detector and the previous one (H-9) should be clarified.
- Page H-12, lines 42-43 – It appears that some of the cost estimates of the instruments have not been updated since Revision 1 (U.S. EPA, 2000). This is a specific example of a potential discrepancy, but there are others. The cost estimates on all the instruments should be updated to 2020 dollars.
- Page H-14, line 1 – Suggest including cadmium telluride (CdTe) detectors. Both CdTe and Cadmium zinc telluride (CZT) detectors are commercially available and utilized in research and Commercial off-the-shelf (COTS) instruments. Are these detectors peltier cooled?
- Page H-14, line 35 – The SAB suggests a brief discussion on segmented detectors.
- Page H-14, lines 36-37 – Clarification should be added to these costs to indicate that this is the cost of the detector without the associated data acquisition system. The MARSSIM document should follow the example for the PIC (Page H-10, lines 34-37). The appendix should strive for consistency in the reporting of costs. This is one example of the discrepancy of reporting. There are other incidents. This becomes more of an issue when presented in Tables H.2 – H.8 where costs are presented for comparison where they are not equivalent.
- Page H-20, line 3 – There is no discussion of neutron detection.
- Page H-20, lines 25-26 – Clarify how increasing flight altitude decreases the Minimum Detectable Activity (MDA). Show the reduction in background contribution for this case

with an infinite areal extent of the contaminant. Highlight differences when the contaminant is localized in a small area.

- Page H-20, lines 23-28 – Replace MDA with MDC.

- Page H-21, lines 27-31 – Some of these TLD materials are not listed in Chapter 6. The SAB suggests that these lists be consistent.

- Page H-28, line 3 – Suggest adding a definition of thoron. Recommend differentiating thoron from radon.

- Page H-33, line 7 – The equipment for the laboratory is commercially available, so is the reference to the equipment being in the "testing phase" referring to the fieldable equipment? Clarification is needed.

- Page H-33, line 10 – Suggest adding a definition for "nondestructive." Laser ablation is destructive to the surface which is being ablated.

- Page H-34, lines 8-9 Activity concentration values for the given mass of U or Th appear to be incorrect. For 1 ppm, the ^{238}U activity concentration is 12.3 Bq/kg (assuming pure ^{238}U and not natural uranium or some other combination of uranium isotopes) and the ^{232}Th activity concentration is 4.05 Bq/kg.

- Page H-40, line 1 – Why repeated under "Beta Particle Analysis" for Laboratory equipment. Why not use the same scheme as was done in the description of the field equipment?

- Page H-42, line 3 – Alpha should be listed as a primary mode of detection, not secondary.

- Page H-43, line 38 – Suggest rephrasing "not totally linear" for the energy calibration curve. Non-linearity is primarily manifested in the detector energy resolution.

- Page H-46, line 21 – Suggest clarifying what "a reasonable price" is. What is reasonable to one reader may not be reasonable to another.

- Page H-50, line 12 – Suggest updating the data to 2020. There is/was a facility at the University of Georgia which was thought to be in place before 2012. Consider deleting the reference to the number and where they are located.

- Page H-53, line 16 – It is not a perfect vacuum, so "all" air molecules have not been pumped out.

- Page H-56, lines 5-6 – The instrument is commercially available. Not clear why Approximate Cost is "not available".

- Page H-57, line 2 – Suggest either adding a discussion of field usage or removing the reference to it.

- Page H-59, Table H.2, column headers, 5th column – Update costs to 2020. Some of these values seem to indicate only the cost of the probe, while others include the probe and the instrument to read out the signal. Would be good to present consistently, but at least clarify if it is just the cost of the probe.

- Page H-61, Table H.2, 2nd row – Clarification into what "low-resolution" spectrometers are being referred to would help the reader. There is an indication that sample is under vacuum, which indicates the best available energy resolution, assuming everything is operating properly.

- Page H-62, Table H.3, 4th row, 4th column – Should it read as The LSC process is highly selective …". This is interesting since Alpha radiation was previously indicated (Page H-42, line 3) to be a secondary radiation for alpha detection.

- Appendix H – It seems appropriate to include "Direct Ion Storage" (DIS) devices in Appendix H. TLDs, OSLs, Electronic Dosimeters (EDs) are all presented, but DIS is not. DIS devices are commercially available and are "drop in" replacements for TLD and OSL.

- Pages H-59 – H-72, Tables H.2-H.8 – To the extent practicable and appropriate for the purpose of implementing guidance, list performance specifications for each instrument. These can include efficiency, energy range, energy linearity, resolution, signal to noise ratio, spectral deconvolution, dead time, specialized operator/radiochemistry training, transmission efficiency, isotopic fractionation effects, mass linearity, resolution, and sensitivity.

- Pages H-59 – H-72, Tables H.2-H.8 – Information in Table H.1 is not consistent with that in Table H.7.

Additions to the Radon Bibliography:

ANSI/AARST (American National Standards Institute/ American Association Radon Scientists Technologists). 2014. Protocol for Conducting Measurements of Radon and Radon Decay Products In Schools and Large Buildings, ANSI/AARST MALB-2014, 34 p. https://standards.aarst.org/MALB-2014/index.html

ANSI/AARST (American National Standards Institute/ American Association Radon Scientists Technologists). 2015. Performance Specifications for Instrumentation Systems Designed to

Measure Radon Gas in Air, ANSI/AARST MS-PC-2015, 24 p. https://standards.aarst.org/MS-PC-2015/index.html

ANSI/AARST (American National Standards Institute/ American Association Radon Scientists Technologists). 2019. Radon Measurement Systems Quality Assurance, ANSI/AARST MS-QA-2019, 24 p. https://standards.aarst.org/MS-QA-2019/index.html

ANSI/AARST (American National Standards Institute/ American Association Radon Scientists Technologists). 2020. Protocol for the Collection, Transfer and Measurement of Radon in Water, ANSI/AARST MW-RN-2020, 48 p. https://standards.aarst.org/MW-RN-2020/index.html

Baskaran, M. 2016. Radon Measurement Techniques, *in*: Baskaran, M. (Ed.), Radon: A tracer of geological, geophysical and geochemical studies. Springer International Publishing, Switzerland, pp. 15-35. https://standards.aarst.org/MW-RN-2020/index.html

IEC (International Electrotechnical Commission). 2006. Radiation Protection Instrumentation - Radon and Radon Decay Product measuring Instruments - Part 1: General principles, IEC 61577-1, International Electrotechnical Commission, Geneva, Switzerland. https://standards.globalspec.com/std/380787/IEC%2061577-1

IEC (International Electrotechnical Commission). 2000 Radiation Protection Instrumentation - Radon and Radon Decay Product Measuring Instruments - Part 2: Specific Requirements for Radon Measuring Instruments, IEC 61577-2, International Electrotechnical Commission, Geneva, Switzerland. https://standards.globalspec.com/std/1694803/IEC%2061577-2.

IEC (International Electrotechnical Commission). 2014. Radiation protection instrumentation - Radon and radon decay product measuring instruments - Part 3: Specific requirements for radon decay product measuring instruments, IEC 61577-3, International Electrotechnical Commission, Geneva, Switzerland. https://standards.globalspec.com/std/9897537/ds-en-61577-3

ISO (International Organization for Standardization). 2019. ISO 11665-1:2019 - Measurement of radioactivity in the environment -- Air: radon-222, p. 33. https://standards.iteh.ai/catalog/standards/iso/950298e5-4976-418a-9b5b-d8d4aa7ad47e/iso-11665-1-2019

Subba, Ramu M.C., Raghavayya, M., Paul, A.C. 1994. Methods for the measurement of radon, thoron and their progeny in dwellings, *AERB Technical Manual*, TM/RM – 1

Vaupotič, J, Smrekar, N, Žunić, Z.S. 2017. Comparison of radon doses based on different radon monitoring approaches: *Journal of Environmental Radioactivity*, V. 169–170, P. 19-26. https://doi.org/10.1016/j.jenvrad.2016.11.023

Vyletělová, P, Froňka, A. 2019. Continuous radon-in-water monitoring—comparison of methods under laboratory conditions and results of in situ measurements*: Radiation Protection Dosimetry,* V. 186(2-3), P. 406–412. https://doi.org/10.1093/rpd/ncz241

WHO (World Health Organization) 2009. Radon measurements: *in* editors, Hajo Zeeb, and Ferid Shannoun, *WHO handbook on indoor radon: a public health perspective*, Geneva, Switzerland, p. 21-40.

2.2.2. Charge Question 2.2. Ranked Set Sampling

Please comment on whether the additional optional methodology for the use of Ranked Set Sampling (Appendix E) for hard-to-detect radionuclides is useful, appropriate and clear.

Background on Ranked Set Sampling

Before answering the question as stated, it would be useful to give a brief non-mathematical description of what ranked set sampling is and under what circumstances it might be considered useful for assessing the status of a site contaminated with radioactive material.

Ranked set sampling is a procedure for improving the efficiency of a random sampling design in cases where the main substance of interest is hard to measure precisely (for example, a hard-to-detect radionuclide), but there is some alternative "easy-to-measure attribute" (e.g., soil particle size) which is highly correlated with the substance of interest. The idea is to use the easy-to-measure attribute to rank the sites by their projected level of contamination, and thereby to select a subset of sites to measure the hard-to-detect substance, in the hope that this will give a more representative distribution of measurements than would be achieved by simple random sampling (SRS).

To give more idea what this really involves, here is one example of how the method might be applied. Select nine potential sampling sites and measure the easy-to-measure attribute at each of these. Divide the sites randomly into three groups of three and rank the observations within each group. From the first group, select the site with the highest contamination level (as determined by the easy-to-measure attribute), the site with the middle level in the second group, and then site with the lowest level in the third group. Then, measure the hard-to-detect radionuclide at each of the three selected sites. This gives an RSS of size three. Repeat as many times as needed, e.g., if the sampling scheme as determined by the Measurement Quality Objectives requires a total of 15 measurements, the procedure just described will be repeated five times.

There are of course many variants on this basic idea. In particular, the same procedure could just as easily be applied based on four groups of size four, or five groups of size five, or in principle, groups of any size, but as far as we are aware, the method is not applied in practice for groups of size larger than five.

The mathematical theory of this method shows that the RSS estimator of a population mean is unbiased, and the variance will never exceed that of the SRS, even if the two attributes are poorly correlated. However, the method naturally works best when the two attributes are highly correlated, and there is no clear-cut answer to how highly correlated they need to be. It would be highly desirable if the report included explicit discussion of these issues – how should one assess, in practice, whether the conditions for RSS are met?

It's possible to calculate a theoretical variance for the RSS estimator if the ranking is essentially perfect (i.e., correctly identifies the smallest, middle, and largest values in each block – this doesn't require perfect correlation between the two measurements but would need something close to that).

The numbers in Tables E.1-E.4 do assume perfect ranking. Also, it's not stated explicitly, but the panel believes there's an implicit normal distribution assumption in Tables E.1-E.3.

Preliminary assessment of the RSS method by the SAB leaves some doubts about the frequency that the RSS method would be of genuine help to refine estimates of the activity levels of a hard to detect radionuclide. The one realistic example provided seems to be rather unique, i.e., a pure beta emitter (^{99}Tc) measured either by field measurements of beta activity (easy to measure), or by radiochemical analysis in the laboratory (expensive to measure).

Significant information and guidance on how to perform this type of assessment is missing and (confusingly) a crucial reference (Vitkus, 2012) does not even mention RSS sampling nor is the dataset used in the example mentioned in the paper. Upon some investigation, a YouTube video[5] produced by ORAU was found that seems to discuss the analysis in question for ^{99}Tc. It seems to the SAB that a very strong correlation between the hard to measure and easy to measure radionuclide needs to exist before this method will show important gains in evaluating the mean (or median) concentrations over that of the usual simple random sampling (SRS) computations. Unfortunately, the level of this correlation in the Vitkus data is not described.

A better discussion of practical issues is recommended, such as prerequisites for using a specific method, suitable alternatives (e.g., stratified sampling, regression estimator, ratio estimator), how the easy-to-measure attribute should be determined, and necessary correlations for the method to be useful. Another suggestion was to use calibration methods to assess the relationship between the two attributes as a possible alternative to RSS sampling, which could be similar to regression or ratio estimation.

Moreover, the draft MARSSIM document does not discuss whether the assumptions are likely to be satisfied for radiological survey applications – for example, data consisting of counts may not be well approximated by normal distributions, and it's not clear whether an easy-to-measure attribute exists that has the desired properties.

The SAB recommends that the discussion be extended to other kinds of decision-based outcomes, beyond quoting a mean and its standard error. If this method improves on SRS for calculating confidence intervals, it should improve it for other kinds of statistical outputs as well, but that isn't brought out in the discussion.

Finally, there should be some cautionary text about when not to use this method.

5 https://www.youtube.com/watch?v=qWaqt0E2jeU

Comments on the Examples

The text includes five worked examples. These are somewhat critical to the assessment of this appendix, given that the panel's overall concerns are much more about the practical applicability of the ranked set sampling methodology than the theory behind it. The rationale, theory, assumptions, and limitations should be integrated with the tools to be used in MARSSIM, Revision 2.

However, Examples 2 to 4 are primarily illustrations of the numerical calculations rather than real-data examples, so the review panel's primary focus was on examples 1 and 5. Example 1 elicited a mixed reaction. One reviewer thought it was a well-constructed example that very nicely illustrated the main principles behind ranked set sampling. Another reviewer, however, was critical because the example did not explain numerous practical details, such as stating which radionuclides were being sampled and by what instruments, how the authors accounted for uncertainty, and other important features of the experimental design.

Close examination of data shows that this is a very unusual dataset. There are 12 sites for which both measurements are given (the laboratory analyses and the field screening measurements). The ordering of those 12 sites is exactly the same by both the laboratory analyses and the field screening measurements.

Even though this is the aim of ranked set sampling, it is surely unusual that this aim is exactly achieved in a practical application, and there is no discussion of what the experimenter should do when this is not achieved. Moreover, the sample correlation coefficient of the two sets of measurements is 0.998. This is unrealistically high and indeed raises the issue of why anyone would perform the laboratory analyses, if field screening measurements exist that are so highly correlated. One could simply apply whatever linear regression function best predicts the laboratory analyses from the field screening measurements and treat that as if it were exactly representative of the laboratory analyses.

The suspicion this creates is that this is just a made-up example, not even based on any real dataset. If this is true, that might still be acceptable for the purposes of illustrating the sampling method, but the report should acknowledge that fact and make clear that real data sets are extremely unlikely to display such high correlations.

These concerns prompted the reviewers to take a closer look at Example 5. This is a clearer example with a lot of specific experimental detail, but it still raised concerns where the example came from. The cited source is:

Vitkus, T.J. 2012. Technical Bases and Guidance for the Use of Composite Soil Sampling for Demonstrating Compliance with Radiological Release Criteria. Oak Ridge Institute for Science and Education, Oak Ridge, TN

Vitkus' paper is about composite sampling, not ranked set sampling – another technique that the MARSSIM writing group might have chosen to highlight, but they didn't except for citing this one reference (and one tangential comment in a different part of the draft manual). So, while a lot of the experimental detail seems to have been taken from Vitkus' paper, the discussion of ranked set sampling is not. Moreover, the one specific data example given here (the table on page E-23) does not appear anywhere in Vitkus' paper. The YouTube video mentioned above provides some information about the methodology used in this example (i.e., for the field and laboratory analyses) but this falls short of providing all the information needed to evaluate the method.

These issues should be clarified: is the Vitkus (2012) paper even the right citation, or if it is, what exactly is the source of the data? The correlation between the easy-to-measure beta counts and the results of the (more expensive) laboratory analyses to directly measure ^{99}Tc in soil should be provided, if this correlation is high then Example 5 would seem to provide a good rationale for using RRS over SRS at least for ^{99}Tc. More generally, MARSSIM Revision 2 should give more information about how often one expects to see easy-to-measure and hard-to-measure pairs of radionuclides (or cheap/expensive methods to measure the same radionuclide) with a high enough correlation to make the RSS a commonly used alternative sampling technique.

Based on these comments, the SAB is uncertain about the practical applicability of the RSS method for MARSSIM Revision 2; while the method is widely known in some fields (e.g., ecology), the examples given in the draft MARSSIM only provide a limited basis on which to judge the general applicability of the method for radiological cleanup activities. More discussion of the key points (existence of paired methods with both a good correlation and a significant cost difference between them) is needed. The information provided in the document only marginally increases the confidence of the SAB that this method is helpful enough to be included in the MARSSIM Revision 2. If these issues cannot be clarified, the SAB suggests removing Appendix E.

Additional Comments:

- Page E-2: reference to software produced by the Pacific Northwest National Laboratory (PNNL). It would be helpful to clarify exactly what this software is and for what purpose it is being recommended (there does not seem to be any later reference to PNNL – was this just "mentioned in passing" or does EPA intend that use of this software would be a major part of the recommendation?).

- Page E-6 and subsequently: The authors focus on the sign test as the main test they are recommending, while referring to Chen's book (*Ranked Set Sampling: Theory and Applications* by Zehua Chen, Zhidong Bai and Bimal Sinha, published by Springer in 2004) for the alternative Wilcoxon-Mann-Whitney (WMW) procedure. This is acceptable so long as they are not expecting the WMW method to be widely used – otherwise include an explicit section about that as well.

- Regarding the assumptions behind Tables E.1 – E.3, most likely two things are missing that should be clearly stated. Do the calculations rely implicitly on the assumption of a normal distribution? If so, it should be clearly stated. If the fit to a normal distribution is improved by making some transformation (e.g., power law or logarithmic), it might be beneficial to assess Δ and σ on those scales as well. Normal distributions are not required for the sign test. However, if the assessment of power assumes normal distributions that may affect application of the test, those points should be added to the discussion. Also, the description and formulas on page E-10 appear to communicate that the ranking is perfect (i.e., correctly identification of the elements as smallest, middle and largest among the ranked samples). In practice, the ranking won't be perfect, and there should be some acknowledgement of that fact.

- Section E.2.3: In contrast to Section E.2.2 that derives power from theoretical calculations, Section E.2.3 seems to recommend that the RSS sample be analyzed as if it were a simple random sample (SRS), which will, in particular, overestimate the population variance. It is argued that this is favorable because the test will be more conservative. Because that logic contradicts the purpose of performing an exact power calculation, further discussion of the rationale is recommended.

- Tables E.4 and E.5 should be assessed to determine how much they are really different in practice. Table E.4 leads to smaller critical values but these changers are marginal in many cases. This could be of practical importance in deciding when to use RSS, which presumably does involve some increased costs and therefore should not automatically be recommended.

- Section E.3 and Example 5: do the authors have any comments on the possible effect of spatial dependence on these conclusions? Spatial dependence (autocorrelation) refers to clustering in the data due to various reasons (for example underlying trends in the soil). The theory of ranked set sampling mostly ignores this possibility and assumes that observations are independent. In specific contexts (like Example 5 in Appendix E) it might be possible to test for the presence of autocorrelation in the measurements over a site. The potential effect of spatial dependence on the power or type I error rate of the RSS method should be considered, and if possible include some brief comment on this. Regardless of employing SRS on page E-20 or the more complicated RSS on page E-22, clarification of the reliance on statistical independence of the measurements as an implicit assumption needed for inference about mean concentration levels is needed, including its implications on sampling recommendations.

- In Step 12 of the Example 5 analysis details, it states the following:

  > "The net result is that 15 laboratory samples were required for the SRS sign test, but the requirements were increased to 45 laboratory samples to account for areas of elevated concentration of radioactive material that could reasonably be expected. This process closely parallels the more familiar required/actual scan MDC paradigm used successfully for MARSSIM soil surveys involving gamma-emitting radionuclides"

  To make this example more understandable, how one reaches the new conclusion (45 required samples versus 15 required samples) should be explained.

Summary and Conclusions:

- Usefulness of the method in the radiometric testing context should be better described.

- More thorough discussion in support of examples is recommended, including method selection criteria and how to use the results for decision making.

- Alternative methods, such as regression or calibration, should be added.

Overall, the SAB identified specific shortcomings of the as-written Appendix E. The method is mathematically sound, but its description lacks attention to important practical details. Given the overall length and complexity of the MARSSIM document, the basis for including ranked set sampling should be strengthened. If not, consideration should be given to dropping this appendix entirely. The appendix could be replaced by a much shorter section discussing the basic idea of using an easy-to-measure attribute that is correlated with a hard-to-measure radionuclide, the need to know something about the amount of correlation, and the alternatives that might be used with very brief descriptions of how they would be applied, and some related references.

Recommendations:

- Include discussion on how highly correlated the "easy-to-measure" attribute with the "hard-to-detect" radionuclide need to be, and how should one assess whether the conditions for RSS are met.

- Add cautionary text about when not to use the RSS method.

- Extend RSS discussion to other kinds of decision-based outcomes, beyond quoting a mean and its standard error.

- Comparisons should be made with alternative methods, such as regression or calibration.

- There should be more thorough discussion of examples, including whether to use the method at all and how to use the results for other kinds of decision making.

- More details should be given about the source of the data in Example 5, clarification whether the survey area was analyzed using RSS due to existing ^{99}Tc contamination, and if spiked samples were used at point of measurement for demonstration purposes. In addition, how the surveyors verified that the laboratory and field measurements were truly of ^{99}Tc should be explained.

- If these issues cannot be resolved, the SAB recommends removing Appendix E entirely.

- The appendix could be replaced by a much shorter section discussing the basic idea of using an easy-to-measure attribute that is correlated with a hard-to-measure radionuclide, the need to know something about the amount of correlation, and the alternatives that might be used with very brief descriptions of how they would be applied, and some related references

2.2.3. Charge Question 2.3. Examples in Chapter 5

Please comment on whether the new and additional examples provided in Chapter 5 are useful, appropriate and clear?

Overall, the examples are useful and provide a means for users to see how the manual guidance can be implemented for practical situations. However, each example could provide additional detail on calculations and use of the tables cited in the text. It cannot be assumed that the user will be well versed in either sampling design theory or use of the statistical methods utilized in the examples.

The examples that appear in Chapter 5 include:

1. Scoping Survey Checklist, pages 5-7 to 5-9. This checklist is identical to Rev. 1, pages 5-5 to 5-6.

   Note that the section on characterization surveys provides limited guidance on the scoping of contamination in other matrices besides soils (for example, ground water, surface water, and sediments). While such information is important to assessing how site contamination is impacted, MARSSIM does not provide guidance on how to use this additional information to structure site clean-up or future site surveys. The draft MARSSIM document applies only to surface soils and building surfaces. An additional note to this effect would be helpful in this section.

2. Characterization Survey Checklist, pages 5-17 to 5-19. This checklist is identical to Rev. 1, pages 5-16 to 5-17.

3. Remedial Action Support Survey Checklist, page 5-21 is identical to Rev. 1, page 5-20. Starting with the same section on final status survey (FSS), the two documents provide descriptions and examples that diverge significantly. Both contain flow charts that illustrate the process of a FSS, including identifying measurement locations and data needs for assessment of potential areas of elevated activity. However, draft MARSSIM document adds a flow chart for designing an integrated survey plan that starts with selecting the appropriate scenario (Scenario A or Scenario B) as the basis of the survey design. The draft MARSSIM document also goes into significant details about selecting the appropriate scenario to demonstrate that residual radioactive material levels meet the release criteria and provides an expanded discussion of the gray region and the selection of a lower bound of the gray region (LBGR). Scenario B is a new approach added to Revision 2.

4. Use of WRS Test under Scenario A, page 5-31. This example is modified from the example in MARSSIM Revision 1(U.S. EPA, 2000) that starts on page 5-29, both examples describe how to obtain the number of data points needed to demonstrate compliance

under Scenario A using the WRS test. The description of the process for selecting the number of samples in the reference and test area is clearer in the draft MARSSIM document than in Revision 1.

5. Use of WRS Test under Scenario B, page 5-33 is a new example.

   The SAB recommends including a cautionary note about when not to use the WRS in the example in Chapter 5. The WRS is discussed in Section 8.4.1 (page 8-27). Users of the MARSSIM manual may have not seen this guidance before applying site survey design in Chapter 5.

6. Use of Sign Test under Scenario A, on page 5-34, provides another example for a Scenario A decision on the number of samples to be selected. It is modified from a similar example in Revision 1 starting on page 5-33 and illustrates the use of the Sign Test to obtain the number of data points needed from the survey unit. The example clearly identifies the input decisions needed by the project team on Type I and II decision errors, the $DCGL_W$ and the LBGR value. It also clearly shows the calculation involved and the use of the Sign Test table. This relates directly to the Type I and Type II decision error rates that are based on the project DQO values.

   As in the previous example, a cautionary note about when not to use the Sign Test would be helpful. The Sign Test is addressed in Section 8.3 (page 8-19). The reader may not seek this guidance before applying effort in Chapter 5.

7. Determination Whether Additional Data Points are Required, pages 5-39 and 5-40. This example is modified from Revision 1, page 5-39.

   The Examples #7 and #8 provided in Section 5.3.5 demonstrate how to use the information discussed in Section 5.3.5. However, there are assumptions made for the initial calculation of the number of samples to be taken that should be stated in the example so that the process for determining the numbers is clear to the reader. For instance, in Example 7, it is not obvious where the value of 27 for the number of data points needed for statistical testing came from. There is a need to show this calculation and the use of Table 5.3 in the example.

8. Determination Whether Additional Data Points are Required, pages 5-40 to 5-42 is a modified example from Revision 1 (U.S. EPA, 2000), page 5-39.

   Example 8 provides good contrast to the scenario depicted in Example 7 and allows the user to see how MARSSIM can be an iterative process. A minor item that should be corrected in Example 8 on pg 5-41 is the following:

   > "The grid area encompassed by a triangular sampling pattern of 10 m is approximately ~~86.6~~ m^2, as calculated using Equation 5-3:" 86.6 m^2 should be corrected into 99.7 m^2.

9. Random Sampling Pattern, page 5-45 is identical to the example in Revision 1 (U.S. EPA, 2000), page 5-41.

   Example 9 is used to demonstrate a random sampling pattern for a Class 3 area. It incorrectly references Table I.11 as it should be Table I.12. This type of example is beneficial to the user providing guidance on random selection of the sampling locations. However, it is unclear how Table I.12 is used to select the locations depicted on the diagram. The method used should be explained in some detail.

10. Illustration of a Triangular Systematic Pattern in an Outdoor Class 2 Survey Unit, pages 5-46 to 5-47. This example is identical to the example in Revision 1 (U.S. EPA, 2000), page 5-43.

    Example 10 provides a good means of showing the triangular grid sampling pattern based on the equations used in the chapter. The random start coordinates are stated as a foregone fact without any development. The stated location relates to Table I.12 but it is not obvious how a user identifies that random start point. Some additional guidance on the determination of the random start point would be helpful.

11. Final Status Survey Checklist, pages 5-56 to 5-58 is modified from Revision 1 (U.S. EPA, 2000), pages 5-53 to 5-55.

    Example 11 identifies a sample checklist for the FSS. It provides additional details not included in the Revision 1 version. Checklists such as these provide users with both goals to attain during the remediation and final survey as well as serving as a final quality check that all items have been considered.

Of the 11 examples that appear in Chapter 5, five are unchanged, five are modified (with additional information), and one is new.

The new example #5 describes Scenario B and is a necessary addition to MARSSIM. The unchanged examples were mostly reformatted such that they are easier to read and understand. The modified examples included some additional information or steps or changes to data such that they are better suited to inform the reader. For some of the examples, additional information and guidance would be helpful to the user. In particular, there are some examples that should have additional references to other sections of the MARSSIM document.

In summary, the examples in Chapter 5 are a marked improvement over examples in MARSSIM Revision 1, but additional detail would be useful for better clarity.

2.3. Charge Question 3. Presentation of Information

Is the information in MARSSIM, Revision 2 clear, understandable and presented in a logical sequence? How can the presentation and content of material be modified to improve the understandability of the manual?

Overall, the information in the draft MARSSIM, Revision 2 is reasonably clear and understandable. However, not all information is presented in a logical sequence. The presentation and content of materials can be modified to improve the understandability of the manual in many parts of the draft document. There is a considerable learning curve encountered by even a statistically well-trained reader of MARSSIM which is partly due to the very specialized terminology used to communicate necessarily complex ideas throughout the MAR series of manuals. The response to Charge Question 1.2 makes a number of suggestions designed to assist the novice user to come up to speed more rapidly.

First, the revised description in the manual on how to set the Lower Bound of the Gray Region (LBGR) tried to convey the point that LBGR should be set using site-specific information about the remaining residual contamination rather than some rule of thumb of a more general nature. The manual addresses this point by emphasizing that "For Scenario A, the LBGR is typically chosen to represent a conservative (slightly higher) estimate of the mean concentration of residual radioactive material remaining in the survey unit at the beginning of the FSS. If there is no information with which to estimate the residual radioactive material concentration remaining, the LBGR may be initially set to equal one-half of the $DCGL_W$".

The SAB has identified some ways in which the presentation could be made more understandable and thus more cogent. It should indicate clearly in Section 5.3.3.1 that in Scenario A, the LBGR is the discrimination limit (DL), which is the level of radioactivity below which the measurement is statistically considered to be below the $DCGL_W$, which is the UBGR.

No explanation is given in the manual as to how the suggested rule for setting the LBGR to be 50% of the $DCGL_W$ was arrived at. The LBGR is given by the UBGR in combination with the acceptable error rate for declaring that the amount remaining in the site is below the $DCGL_W$ when in fact it is at or above the $DCGL_W$, i.e., the error rate is the probability that the survey result comes in below the LBGR when the true concentration of radioactivity in the site is at or above the $DCGL_W$. A conservative estimate of the amount thought to remain in the site was chosen as a recommended LBGR because it would allow the fewest samples/least expensive testing method (largest σ) without having a LBGR that is likely to be frequently exceeded when the amount remaining in the site is actually less than the $DCGL_W$.

The SAB has suggested another possible rule of thumb (see Section 2.3.1) and shown the derivation of the suggested rule for setting the LBGR to be 50% of the $DCGL_W$. The SAB suggests adding some pertinent graphics illustrating the concepts involved with distribution plots of the measurement results expected under the null and alternative hypotheses. In addition, the SAB has pointed out that there are situations in which it is simply unavoidable to use a rule of thumb.

Second, the SAB finds the presented information in the draft MARSSIM document, with regard to the use of the term "area factor" to be confusing. As noted in the response to Charge Question 3.2, improvements could be made to fully remove the use of the "area factor" designation and simply utilize the ratio of the Elevated Measurement Comparison (EMC) release criteria to the wide-area release criteria, and to emphasize $DCGL_{EMC}$ should be based on site-specific modeling or calculations.

Third, with regard to the effectiveness of the new organization of Chapter 4, the SAB finds the information in the draft MARSSIM, Revision 2, Chapter 4 is reasonably clear and understandable, but some improvements could be made, especially for Section 4.8.4.1, where the presented information is inadequate. Moreover, not all information is presented in the most logical sequence, as further discussed in response to Charge Question 3.3.

Fourth, the SAB finds that the changes in the draft MARSSIM document, where certain derivations were moved from Chapter 5 of Revision 1 to Appendix O, are an improvement that makes Chapter 5 clearer and more understandable. However, some additional improvements could be made. Two sections in Appendix O are duplicates of other sections in the draft MARSSIM Revision 2 and should be deleted from Appendix O.

2.3.1. Charge Question 3.1. Lower Bound of the Gray Region (LBGR)

Please comment on the revised description of how to set the Lower Bound of the Gray Region (LBGR) and its likely effectiveness in encouraging users to rely on site-specific information for doing so (Chapter 4 and Section 5.3).

In the case of Scenario A, where the null hypothesis is that the site contains residual radioactivity at or above the allowed $DCGL_W$, the recommendation in the draft manual is to set the Lower Bound of the Gray Region (LBGR) at a *conservative* (slightly higher) *estimate* of the mean concentration of residual radioactive material that is thought to remain in the site. This is emphasized at several locations in the draft manual and is certainly likely to increase the probability that users will use site-specific information in this way, as opposed to a rough rule of thumb based on 50% of the $DCGL_W$.

In practice, however, the value of site-specific information about the residual level of radioactivity is likely to vary from site to site. The suggestion that the LBGR be set to a value near to the median seen in preliminary data is a good one so long as the preliminary data are reasonably informative. In cases where preliminary data are limited, adherence to a heuristic rule (such as using ½ of the $DCGL_W$) probably cannot be avoided.

Another possible rule of thumb that could be added to the draft manual as a suggestion is the LBGR could be based on a gray region that is 1.67 times the standard deviation (σ) of the preliminary measurements, for cases where the available data are very sparse. This is a relative shift of 1.67 ($\Delta/\sigma = 1.67$), which gives a one-sided error rate of 5% for the standard normal distribution, i.e., a gray region that is 1.67σ wide contains 95% of the probability mass of the preliminary measurements if they follow a standard normal distribution.

In the absence of reliable survey data, it may also be reasonable to assume that the standard deviation (σ) of the measurement data is 30% of the $DCGL_W$. This is also referred to as a coefficient of variation (CV) of 30%, where the CV is a measure of the dispersion of the data and is defined by the ratio of the standard deviation σ to the mean, which is taken here to be the $DCGL_W$. A CV of 30% is consistent with the statement in the draft manual that "when preliminary data are not obtained, it may be reasonable to assume a coefficient of variation on the order of 30 percent, based on experience." (Pg 5-29 line 27 – Pg 5-30 line 2).

It may be noted that if both of the two prior suggestions were used in combination, it would equate to using a LBGR that is ½ of the $DCGL_W$, (viz., $\Delta/\sigma = 1.67$, $\sigma/DCGL_W = 0.3$ ➔ $\Delta/DCGL_W = 0.3 \times 1.67 = 0.5$ ➔ $\Delta = 0.5\ DCGL_W$; for Scenario A, $\Delta = DCGL_W - LBGR$; thus, $LBGR = 0.5\ DCGL_W$) as suggested in the draft manual.

The needed sample size can be reevaluated as the data come in and if the median of a first set of observations is far below the $DCGL_W$, the necessary sample size can be lowered. Development of a good case study (as an example) would probably be more persuasive than the very short description of using various possible preliminary data sources that is currently given in the draft manual without giving any specifics. Unless the preliminary data are strong it is natural to fall back on a (however inadequate) "hard and fast" rule.

In Section 5.3.3.1, the manual should indicate clearly that in Scenario A, the LBGR is the discrimination limit (DL) and level of radioactivity below which the measurement is statistically considered to be below the $DCGL_W$, which is the UBGR. It is unfortunate that no explanation is given as to how the suggested rule for setting the LBGR was arrived at. In statistical terms, the LBGR is given by the UBGR in combination with the acceptable error rate for declaring that the amount remaining in the site is below the $DCGL_W$ when in fact it is at or above the $DCGL_W$, i.e., the error rate is the probability that the survey result comes in below the LBGR when the true concentration of radioactivity in the site is at or above the $DCGL_W$. The error rate is determined by the LBGR and the width of the error distribution based on the accuracy of the testing method and sample size (i.e., the σ and n). It is assumed that a conservative estimate of the amount thought to remain in the site was chosen as a recommended LBGR because it would allow fewer samples (n) or less expensive testing method (with larger σ) without having a LBGR that is likely to be frequently exceeded when the amount remaining in the site is actually less than the $DCGL_W$. This is alluded to in the material in Sections 4.12.3.1 and 4.12.3.2 on the Wilcoxon Rank Sum and sign tests. On page 4-56, the LBGR is set at the median from data in a post-remediation survey since it is conservative to use the higher of the mean or median of values. Footnote 9 states "Larger values for the LBGR and σ lead to a smaller relative shift that, in turn, leads to a larger number of required measurements." Also, it is discussed in 5.3.2 that survey designers can consider "increasing the width of the gray region, as long as the LBGR is not set lower than the estimate of the residual radioactive material remaining in the survey unit in Scenario A." Despite these rather cryptic references, the basic concept is not really explained, especially in the part where it is recommended how to set the LBGR, i.e., Section 5.3.3.1. It should be explained at that location in the manual and would serve to convince users that it is a good idea. Furthermore, the detailed material in Sections 4.12.3.1 and 4.12.3.2 could be moved to Section 5.3.3.1 and just referred to at a higher level in Chapter 4.

For Scenario B, the LBGR is the Action Level (AL), and is equal to zero or the $DCGL_W$ (see Section 5.3.4.1). Also, Section 5.3.1 states "individuals designing a MARSSIM Survey using Scenario B should make conservative assumptions for the estimate of the standard deviation (σ) (see Section 5.3.3.2) so that even if the variability in the survey unit is higher than expected, the power of the resulting survey $(1-\beta)$ (see Section 5.3.2) will still be sufficient to ensure that survey units with residual radioactive material in excess of the AL will be discovered $1-\beta$ percent of the time." However, there is no recommendation to evaluate power for various alternative survey methods as part of planning a survey program to clear the site; there is only a requirement for a post hoc power analysis. It would be useful for users planning surveys to evaluate the power of a survey as a function of the DL, the σ and the sample size when negotiating a DL with the regulatory agencies, and some material about this could be added to the manual.

In general, the basic principles involved in Scenarios A and B are not well explained with graphical illustrations anywhere in the draft manual. There are relevant graphics in Appendix D, Section D.1.6.2, one for each scenario, but each one only shows the distribution under the null hypothesis and the α error rate, not the distribution under the alternative hypothesis and the β error rate. It would be much better to have these concepts graphically illustrated in Chapter 5 where the material is presented on the gray regions. Figure 6.1 in Chapter 6 is a nice example of a graphic that illustrates the principles of hypothesis testing with both null and alternative distributions and both types of error rates illustrated. The SAB notes that there is an error in a text box in the figure for Scenario B (Appendix section D.1.6.2), which reads "Survey unit passes if and only if the test statistic falls in the rejection region." Since this is for Scenario B, "passes" should have read "fails" in Figure D.6.

Recommendations:

- The suggestion that the LBGR be set to a value near to the median seen in preliminary data is a good one so long as the preliminary data are reasonably informative.

- When preliminary data are limited the rule of thumb of using ½ the $DCGL_W$ probably can't be avoided. The needed sample size can be reevaluated as the data come in and if the median of a first set of observations is far below the $DCGL_W$ the necessary sample size can be lowered.

- Development of a good case study (as an example) would probably be more persuasive than the very short description of using various possible preliminary data sources without giving any specifics.

- Consider adding a rule of thumb that the LBGR could be based on a gray region that is 1.67 times the σ of the measurements, for cases where the available data are very sparse. Also consider adding a rule of thumb that in the absence of adequate data it may be reasonable to assume that the standard deviation of the measurement data is 30% of the $DCGL_W$. Note that these two assumptions lead to the choice of the LBGR as ½ the $DCGL_W$.

- In Chapter 4, indicate clearly that in Scenario A the LBGR is the DL, which is the level below which the measurement is statistically considered to be below the $DCGL_W$.

- For Scenario B, add some material describing use of a series of power curves as a function of the DL for different sample sizes to evaluate whether the DL for a "reasonable" study size is unrealistically large in relation to typical requirements.

- Add two figures (as in Figure 6.1) to Chapter 4 to illustrate the relationships in Scenarios A and B, and describe in simple language how the relative shift (width of gray area) depends on the values of σ and the α and β error rates.

- Consider moving the material in Section 4.12.3.1 into Chapter 5, where it is also covered, with a higher-level summary of sample size determination being presented in Chapter 4, and reference (as is already given) to the appropriate sections in Chapter 5.

- In Section 5.3.3.1, prior to the examples, include a short paragraph on "Considerations for Setting the LBGR" discussing how the site-specific information along with some emphasis on conservatism is applied and why this was chosen as a basis for setting the LBGR.

2.3.2. Charge Question 3.2. Area Factor
Please comment on whether avoiding the use of the term "area factor" improves understandability of the elevated measurement comparison concept (Section 8.6.1).

Area factor is the ratio of the Elevated Measurement Comparison (EMC) release criteria ($DCGL_{EMC}$) to the wide-area release criteria ($DCGL_W$) and is the magnitude by which the concentration of residual radioactive material in a small area of elevated activity can exceed the $DCGL_W$ while maintaining compliance with the release criteria. It should be based on site-specific modeling or calculations. Due to the misapplication of published area factors from the literature, and to provide focus on the need for development of site-specific EMC criteria, the draft MARSSIM document avoids the use of the term area factor. In addition, lessons learned from training MARSSIM users show that describing the EMC concept in descriptive language, rather than by defining additional terminology, seems to improve understandability of the concept.

However, the term 'area factor' appears to still be utilized in many sections of the draft MARSSIM document. As such, it is not clear if avoiding the use of the term has been complete or would improve the understandability. In fact, the sections that do still refer to an area factor may introduce significant confusion more generally. A few examples from the draft MARSSIM document are discussed below to illustrate potential problems introduced by this attempt:

- Section 5.3.5.1, paragraph on page 5-37, Lines 15-22,

> Revisions 0 and 1 of MARSSIM (published in 1998 and 2000, respectively) included the calculation of an area factor[7] as an intermediate step in the determination of the required scan MDC. The use of an area factor is not necessary if $DCGL_{EMC}$ is tabulated directly as a function of the area of radioactive material. To simplify the determination of the required scan MDC, the use of the area factor as an intermediate calculation is not included in this revision of MARSSIM. The area factor can still be used if the ratio of the $DCGL_{EMC}$ to the $DCGL_W$ is known and will produce the same results as the approach described in the current revision of MARSSIM.

In section 5.3.5.1, *Determination if Additional Data Points are Needed*, a paragraph notes that revisions 0 and 1 of MARSSIM included the calculation of an area factor (note that it is defined, identified as A_m, discussed further in footnotes, is included in the list of symbols, and specifically given much description in Appendix O, Section O.4) as an intermediate step in the determination of the required scan MDC. It then states that the use of an area factor is not necessary if $DCGL_{EMC}$ is tabulated directly as a function of the area of the radioactive material. This statement could be interpreted to mean that an area factor is needed if $DCGL_{EMC}$ is not tabulated in this manner. However, the section further notes that "the use of the area factor as an intermediate calculation is not included in this revision of MARSSIM." But then it further states that the area factor can still be used under certain conditions. This seems patently confusing. The message appears to be, we don't want you to use area factors, but you can if you want to under certain circumstances.

- Section 5.3.8, paragraph on page 5-49, Lines 3-12,

> When an investigation level is exceeded, the first step is to confirm that the initial measurement or sample actually exceeds the particular investigation level. This may involve taking further measurements to determine that the area and level of the elevated residual radioactive material are such that the resulting dose or risk meets the release criteria. Rather than—or in addition to—taking further measurements, the investigation may involve assessing the adequacy of the exposure pathway model used to obtain the DCGLs and area factors, as well as the consistency of the results obtained with the HSA and the scoping, characterization, and RAS surveys. Depending on the results of the investigation actions, the survey unit may require reclassification, remediation, or resurvey. **Table 5.4** illustrates an example of how investigation levels can be developed.

In Section 5.3.8, *Determining Investigation Levels*, a paragraph discussing steps to take when an investigation level is exceeded, points out that rather than – or in addition to – taking further measurements, the investigation may involve assessing the adequacy of the exposure pathway model used to obtain the DCGLs and *area factors*, etc. This is confusing, as Section 5.3.5.1 has already noted that area factors are not necessary for Revision 2. The SAB recommends deleting the reference to area factors here.

- Section 8.6.1. Elevated Measurement Comparison.
The ratio of the Elevated Measurement Comparison (EMC) release criteria ($DCGL_{EMC}$) to the wide-area release criteria ($DCGL_W$) is the magnitude by which the concentration of residual radioactive material in a small area of elevated activity can exceed the $DCGL_W$ while maintaining compliance with the release criteria. This ratio was previously referred to as the "area factor." Section 8.6.1 discusses the related concept of residual material at the $DCGL_{EMC}$ over a small area being equivalent to residual material at the $DCGL_W$ over a large area and while being technically sound. Section 8.6.1 does not make this point clearly.

In addition, the point that derivation of the $DCGL_{EMC}$ is beyond the scope of MARSSIM, and strictly determined by regulatory agencies is not included in this section either. This important information from Section 4.5.2 should be reiterated in Section 8.6.1.

- Section 8.6.3, Example 14 box on page 8-49.

Example 14: Class 1 Survey Unit with Elevated Areas

Consider a Class 1 Survey unit that passes the nonparametric statistical tests and contains some areas that were flagged for investigation during scanning. Further investigation, sampling, and analysis indicate one area is truly elevated. This area has a concentration that exceeds the derived concentration guideline level determined using the Wilcoxon Rank Sum test by a factor greater than the area factor calculated for its actual size. This area is then remediated. Remediation control sampling shows that the residual radioactive material was removed, and no other areas were affected by residual radioactive material. In this case, one may simply document the original final status survey (FSS), the fact that remediation was performed, the results of the remedial action support survey, and the additional remediation data. In some cases, additional FSS data may not be needed to meet the release criteria.

This example is predicated on finding one area that is truly elevated, noting that "the area has a concentration that exceeds the derived concentration guideline level determined using the Wilcoxon Rank Sum test by a factor greater than the *area factor* calculated for its actual size." Utilizing the term "area factor" in such an example introduces confusion, as it appears as an important deciding factor in describing if an area is truly elevated. However, Section 5.3.5.1 has already noted that area factors are not necessary for MARSSIM Revision 2. This example should be re-written to remove this confusion.

- Appendix E, the table for Example 5, referring to ^{99}Tc $DCGL_{EMC}$ Information on page E-10.

^{99}Tc $DCGL_{EMC}$ Information					
Area Factors	10 m^2	20 m^2	50 m^2	100 m^2	200 m^2
	74.6	43.9	21.1	11.6	6.2
$DCGL_{EMC}$ (Bq/kg)	54,100	31,800	15,300	8,400	4,480
$DCGL_{EMC}$ (pCi/g)	1,462	860	413	227	121

Abbreviations: 99Tc = technetium-99; DCGL = Derived Concentration Guideline Level; EMC = elevated measurement comparison; Bq/kg = becquerels per kilogram; pCi/g = picocuries per gram.

This table, and associated discussion, is provided in an example of a survey design for hard-to-detect radionuclides using ranked set sampling, as illustrated for technetium-99 (^{99}Tc). Survey design parameters are provided for the example in both text and tabular form. These specifically include "area factor" designations. The SAB recommends deleting the reference to area factors here.

- Appendix O, Section O.4 on pages O-4 through O-8.

Appendix O.4 otherwise is called out in only one place in Section 4.5.3.4 *Gross Activity DCGLs for Multiple Radionuclides in Known Ratios* and is not even referred to in Section 8.6.1. The discussion in Appendix O on area factors is unnecessarily confusing. Recognizing the material in Appendix O is presented as "historical examples," the inclusion of the illustrative examples doesn't seem necessary, unless one were to include examples of how published area factors were misused in the past; such a discussion would make it much clearer why the use of "area factors" has been eliminated.

This entire section discusses 'Calculating Area Factors and the DCGL for the EMC.' While it notes that "Because these area factors were misused for specific problems, the term 'area factor' is largely omitted from the main body of this report," and goes on to say, "Historical information on the use of area factors is provided in this appendix for completeness," this is very confusing to a reader. If this important line is not well understood, users might see the illustrative examples and Table O.4 (Illustrative Examples of Outdoor Area Factors) and Table O.5 (Illustrative Examples of Indoor Area Factors) as able to be used. It is recognized that Appendix O in the draft Revision 2 manual does include footnotes noting these are illustrative only, but these footnotes also direct users to "Consult regulatory guidance to determine *area factors* to be used for compliance demonstration." This returns to the issue in Section 5.3.5.1, which notes that area factors are not necessary. These footnotes seem to say that they are so necessary that users should consult with their regulator. Perhaps, it is better not to include any discussion about area factors if they are not needed. Inclusion of the historical information in the appendix does not appear to be necessary. Further, the summary and conclusions statements in Section O.4.5 note that "It is always acceptable and conservative to assume the smallest area factors possible (i.e., 1)." If area factors are not necessary, then this conclusion is highly confusing to a user.

Perhaps cite some examples of current State Regulator documents (e.g., New Jersey Field Sampling Procedures Manual, NJDEP, 2005) for clarity while clearly noting that the draft MARSSIM document encourages utilizing the ratio of the Elevated Measurement Comparison (EMC) release criteria to the wide-area release criteria.

- Appendices – Glossary, page GL-1, includes a definition of area factor, but does not state that using it is not necessary. Perhaps this should refer to a ratio of $DCGL_{EMC}/DCGL_W$ or not be included in the glossary at all.

area factor (A_m): A factor used to adjust the *derived concentration guideline level ($DCGL_W$)* to estimate the *derived concentration guideline level ($DCGL_{EMC}$)* and the *minimum detectable concentration* for *scanning surveys* in *Class 1 survey* units, wherein the $DCGL_{EMC} = DCGL_W \times A_m$. A_m is the magnitude by which the *concentration* of *residual radioactive material* in a small *area of elevated activity* can exceed the $DCGL_W$ while maintaining compliance with the *release criteria*.

Recommendations:

- Remove the use of the "area factor" designation and simply utilize the ratio of the Elevated Measurement Comparison (EMC) release criteria to the wide-area release criteria, to improve understandability of the EMC concept and emphasize the need for site-specific modeling or calculations.

- Sections that continue to utilize “area factor” (i.e., 5.3.5.1 and 5.3.8) should be revised to remove the ‘historical approach discussions’ if appropriate, simply refer to the ratio of the $DCGL_{EMC}$ to the $DCGL_W$ rather than “area factor,” and delete any reference to “area factor.”

- Section 8.6.3, Example 14 specifically refers to the “area factor” and introduces confusion. Refer to the ratio of $DCGL_{EMC}$ to $DCGL_W$ instead.

- Historical illustrations of area factor in Appendix O should be eliminated.

- Cite some examples of current State Regulator documents for clarity while clearly noting that MARSSIM Revision 2 encourages utilizing the ratio of the Elevated Measurement Comparison (EMC) release criteria to the wide-area release criteria.

- Delete or revise the definition of “area factor” in the Glossary.

2.3.3. Charge Question 3.3. Organization of Chapter 4.

Please comment on the effectiveness of the new organization of Chapter 4 (Considerations for Planning Surveys) to improve the understandability of the Chapter.

The SAB finds the new organization of Chapter 4 to be a considerable improvement over the previous version, and it improves the understandability of the material presented; however, additional work is needed, as described in the following paragraphs.

Section 4.1 states the purpose of Chapter 4 is to “introduce the Multi-Agency Radiation Survey and Site Investigation Manual (MARSSIM) user to general considerations for planning MARSSIM-based surveys,” then in subsequent language narrows the scope to an emphasis on the Final Status Survey. The SAB suggests that the material in Sections 2.4.3-2.4.6, covering all survey types, may more properly belong in Section 4.3, along with Figures 2.4-2.8. At the very least, the SAB recommends this material be referenced in Section 4.3.

Section 4.2 opens with the seven steps of the DQO process; however, the headings in the bulleted paragraphs in Section 4.2.1 are not fully compatible with those steps. Thus, the SAB recommends that each heading mimic the DQO step, as presented in the introduction to Section 4.2. For example, for the first subsection, instead of “Clarify the study objective,” the heading should read, “State the problem (DQO Step 1),” then follow with the explanation of what that means. The SAB also recommends the final paragraph underneath the bulleted list in Section 4.2.1 should be moved into the bulleted list with the heading “Optimize the survey design (DQO Step 7).

Similarly, Section 4.8.2 is on Measurement Quality Objectives (MQOs); however, individual MQOs are not discussed until Section 4.8.4, *Selection of Instruments for Field Measurement*. The heading for Section 4.8.4.1. should read *Method Uncertainty,* the first MQO; however, the heading for this subsection is *Reliability and Robustness*, which causes confusion. The SAB recommends this subsection describe in detail the concepts of method uncertainty and required method uncertainty (see discussion in response to Charge Question 1.3a). This subsection should also include discussion on how to determine the required method uncertainty for every radionuclide of concern; how the required method uncertainty is used in the selection of instruments for survey measurement, and how it is used to determine if the DQOs are met after measurements are made by calculation of measurement uncertainty based on measurement data. Measurement uncertainty is not discussed in this subsection. Its location in the MARSSIM document should be cross-referenced.

The SAB recommends that Figures 4.1 and 4.2 appear in reverse order in the chapter. Figure 4.2 shows the general case, while Figure 4.1 addresses the specific case of the Final Status Survey. The SAB further recommends these two figures be in closer proximity, and suggest incorporating them into Section 4.3, which discusses survey types.

The SAB finds that Section 4.7.1 *Basic Terms*, does not refer the reader to the Glossary. In addition, the Glossary does not include all the terms identified as *Basic Terms*, or there are other inconsistencies. Missing terms include (a) sample median, (b) parametric tests, and (c) Student's t test. For consistency the word "decision" should be included in the term "Type I and Type II errors." The SAB recommends ensuring consistency between the *Basic Terms* enumerated and the Glossary.

The SAB recommends including a discussion in Section 4.8.2 describing the rationale for setting the MDC at less than 50% of the UBGR, since the rationale for this recommendation is not discussed in Chapter 6 to which the reader is referred. Alternatively, the discussion and rationale could be included in Chapter 6.

The SAB finds the removal of information from Chapter 4 to Appendix O improves the readability of Chapter 4, but there is room for additional improvements to be made. The SAB suggests consideration be given to moving Section 4.12 to Appendix O as well as some of the details regarding the calculation of the various types of DCGLs in Section 4.5.3 to further improve the flow of the text. A summary of the DCGLs should be retained in Chapter 4, but the reader referred to Appendix O for the details of the calculations. In addition, whether or not Section 4.12 is moved to Appendix O, Sections O.5 and O.6 should be deleted, as they repeat, word for word, the information in Sections 4.12.8 and 4.12.9, respectively.

The SAB finds there is insufficient notice provided regarding the use of the Unity Rule and recommends that a precautionary statement be added in Section 4.4 to the effect that, "the dose/risk endpoints must be identical and able to be summed in order to use this Eq. 4-3." The SAB also suggests consideration of merging Sections 4.4 and 4.5 because they both discuss use of the Unity Rule.

Recommendations:

- Add the material in Sections 2.4.3-2.4.6, covering all survey types, to Section 4.3, along with Figures 2.4-2.8, or reference this material in Section 4.3.

- Revise the headings in the bulleted paragraphs in Section 4.2.1 to mimic the DQO steps, as presented in the introduction to Section 4.2.

- Move the final paragraph underneath the bulleted list in Section 4.2.1, into the bulleted list above, with the heading "Optimize the survey design (DQO Step 7)".

- Incorporate Figures 4.1 and 4.2 into Section 4.3, reverse their order of appearance, and place them in closer proximity to one another.

- Ensure there is consistency between the *Basic Terms* in Section 4.7.1 and the Glossary.

- The heading for Section 4.8.4.1 should be changed into *Method Uncertainty*. The concepts of method uncertainty and required method uncertainty should be discussed in detail. The determination of required method uncertainty and its use should also be discussed.

- Section 6.4, *Measurement Uncertainty* should be cross-referenced in Chapter 4, Section 4.8.4.1, so that the readers will know where to find this topic.

- Include a discussion in Section 4.8.2 describing the rationale for setting the MDC at less than 50% of the UBGR.

- Move Section 4.12 and details regarding the calculation of the various types of DCGLs in Section 4.5.3 to Appendix O.

- Delete Sections O.5 and O.6.

- Add a precautionary statement in Section 4.4 regarding the appropriate use of the Unity Rule.

- Merge Sections 4.4 and 4.5 because they both discuss use of the Unity Rule.

- As stated in response to Charge Question 3.1, add two figures (as in Figure 6.1) to Chapter 4 to illustrate the relationships in Scenarios A and B, and describe in simple language how the relative shift (width of gray area) depends on the values of σ and the α and β error rates.

- As stated in response to Charge Question 3.1, consider moving the material in Section 4.12.3.1 into Chapter 5, where it is also covered, with a higher-level summary of sample size determination being presented in Chapter 4, and reference (as is already given) to the appropriate sections in Chapter 5.

2.3.4. Charge Question 3.4. Moving Derivations from Chapter 5 to Appendix O

Please comment on the effectiveness of moving derivations from Chapter 5 to Appendix O to improve the understandability of the chapter.

The SAB finds the changes where derivations were moved from Chapter 5 to Appendix O to be an improvement. In MARSSIM Revision 1 (U.S. EPA, 2000), the following tables and accompanying discussions have been removed from the body of Chapter 5 and put into Appendix O of MARSSIM Revision 2. This makes Chapter 5 more readable without losing important background information.

- Table 5.1, Values of P_r for Given Values of the Relative Shift, Δ/σ, When the Radionuclide Is Present in Background. Table 5.1 (MARSSIM Revision 1) has been moved to Appendix O, Table O.1 in Revision 2.
- Table 5.2, Percentiles Represented by Selected Values of α and β. Table 5.2 (Revision 1) has been moved to Appendix O, Table O.2 in Revision 2.
- Table 5.4, Values of P_s for Given Values of the Relative Shift, Δ/σ, When the Radionuclide Is Not Present in Background. Table 5.4 (MARSSIM Revision 1) has been moved to Appendix O, Table O.3 in Revision 2.

The use of the term Area Factor has been discontinued in the draft Revision 2, and the discussion of Area Factor in MARSSIM Revision 1 have been removed (Section 5.5.2.4). Table 5.6, Illustrative Examples of Outdoor Area Factors from Revision 1 has been moved to Appendix O, Table O.4 in Revision 2. Table 5.7, Illustrative Examples of Indoor Area Factors from Revision 1 has been moved to Appendix O, Table O.5 in the draft MARSSIM Revision 2. Descriptive information on Area Factors from Revision 1 has been removed and this information is now in Appendix O.

Section O.5 in Appendix O of the draft MARSSIM document, Release Criteria for Discrete Radioactive Particles, is the same as Section 4.12.8 of Chapter 4 of the same document. Section O.6 in the draft document, Uranium Mill Tailings Radiation Control Act of 1978 (UMTRCA) Sites, is the same as in Section 4.12.9. Since these sections are appropriate for Section 4, Sections O.5 and O.6 should be deleted from Appendix O.

Recommendations:

- Sections O.5 and O.6 should be deleted from Appendix O, as they are duplicates of Section 4.12.8 and Section 4.12.9, respectively.

GLOSSARY

action level (AL): The numerical value that causes a decision maker to choose or accept one of the alternative actions to the "no action" alternative. See also in this glossary *investigation level*.

activity: See in this glossary *radioactivity*.

alpha (α): The specified maximum probability of a *Type I decision error*. In other words, the maximum probability of rejecting the *null hypothesis* when it is true. *Alpha* is also referred to as the *size of the test*. *Alpha* reflects the amount of evidence the decision maker would like to see before abandoning the *null hypothesis*.

alpha particle: A positively charged particle ejected spontaneously from the nucleus of an unstable atom during *radioactive decay* (or disintegration). It is identical to a helium nucleus that has a mass number of 4 and an electrostatic charge of +2. It has low penetrating power and a short range (a few centimeters in air).

alternative hypothesis (H_1): See in this glossary *hypothesis*.

area: A general term referring to any portion of a *site*, up to and including the entire *site*.

area of elevated activity: An *area* over which the *concentration* of *residual radioactive material* exceeds a specified value of the *derived concentration guideline level ($DCGL_{EMC}$)*.

becquerel (Bq): The International System (SI) unit of *activity* equal to one nuclear transformation (disintegration) per second. 1 Bq = 2.7×10^{-11} curies (Ci) = 27.03 picocuries (pCi).

beta (β): The probability of a *Type II decision error* (i.e., the probability of accepting the *null hypothesis* when it is false). The complement of *beta* $(1-\beta)$ is referred to as the *power* of the test.

beta particle: A charged particle (with a mass equal to 1/1,837 that of a proton) that is emitted from the nucleus of an unstable atom during *radioactive decay* (or disintegration). A negatively charged *beta particle* is identical to an electron, while a positively charged *beta particle* is called a positron.

bias: The *bias* of a *measurement method* is a persistent deviation of the *mean* measured result from the true or accepted reference value of the quantity being measured, which does not vary if a *measurement* is repeated.

calibration: The set of operations that establish, under specified conditions, the relationship between values indicated by a measuring instrument or measuring system, or values represented by a material measure, and the corresponding known value of a measurand.

Class 1 area: *Areas* that have, or had prior to *remediation*, a potential for *residual radioactive material* (based on *site* operating history) or known *residual radioactive material* (based on previous *radiation surveys*) above the *derived concentration guideline level (DCGL$_W$)*. Examples of *Class 1 areas* include: (1) *site areas* previously subjected to *remedial action*s, (2) locations where leaks or spills are known to have occurred, (3) former burial or disposal *sites*, (4) waste storage *sites*, and (5) *areas* with *residual radioactive material* in discrete solid pieces of material and high specific *activity*.

Class 2 area: *Areas* that have, or had prior to *remediation*, a potential for *residual radioactive material* or known *residual radioactive material*, but are not expected to exceed the *derived concentration guideline level (DCGL*$_W$). To justify changing an area's *classification* from Class 1 to Class 2, the existing data (from the *Historical Site Assessment [HSA]*, *scoping surveys*, or *characterization surveys*) should provide a high degree of confidence that no individual measurement would exceed the *derived concentration guideline level (DCGL$_W$)*. Other justifications for this change in an area's classification may be appropriate based on the outcome of the *Data Quality Objectives (DQO) process*. Examples of *areas* that might be classified as Class 2 for the *final status survey* include: (1) locations where radioactive materials were present in an unsealed form (e.g., process facilities), (2) transport routes with potential *residual radioactive material*, (3) *areas* downwind from stack release points, (4) upper walls, roof support frameworks, and ceilings of buildings or rooms subjected to airborne radioactive material, (5) *areas* where low concentrations of radioactive materials were handled, and (6) *areas* on the perimeter of former radiological control *areas*.

Class 3 area: Any *impacted areas* that are not expected to contain any *residual radioactive material* or are expected to contain *concentrations* of *residual radioactive material* at a small fraction of the *derived concentration guideline level (DCGL*$_W$), based on *site* operating history and previous *radiation surveys*. To justify changing an area's *classification* from Class 1 or Class 2 to Class 3, the existing data (from the *Historical Site Assessment [HSA]*, *scoping surveys*, or *characterization surveys*) should provide a high degree of confidence that there is either no *residual radioactive material* or that any levels of *residual radioactive material* are a small fraction of the *DCGL$_W$*. Other justifications for this change in an area's classification may be appropriate based on the outcome of the *Data Quality Objectives (DQO) process*. Examples of *areas* that might be classified as Class 3 include buffer zones around *Class 1* or *Class 2 areas*, and *areas* with very low potential for *residual radioactive material* but insufficient information to justify a *non-impacted classification*.

cleanup standard: A numerical limit set by a regulatory agency as a requirement for releasing a *site* after *cleanup*. See in this glossary *release criteria*.

coefficient of variation: A unitless measure that allows the comparison of dispersion across several sets of data. It is often used in environmental applications because variability (expressed as a *standard deviation*) is often proportional to the *mean*. The *coefficient of variation* of a nonnegative random variable is the ratio of its *standard deviation* to its *mean*.

commutability: Commutability is defined as property of a given reference material, demonstrated by the closeness of agreement between the relation among the measurement results for a stated quantity in this material, obtained according to two measurement procedures, and the relation obtained among the measurement results for other specified materials.

NOTES 1 The material in question is usually a calibrator. 2 At least one of the two given measurement procedures is usually a high-level measurement procedure.

confidence interval: An estimated range of values for which there is a specified probability (*e.g.*, 80%, 90%, 95%) that this range contains the true value of an estimated parameter, such as the true mean, the estimated range being calculated from a given set of *sample* data.

critical level (L_C): The level at which there is a statistical probability (with a predetermined confidence) of correctly identifying a *measurement* as greater than background.

critical value: A fixed value of the *test statistic* corresponding to a given probability level, as determined from the probability distribution of the *test statistic*. The value of a statistic (t) corresponding to a given significance level as determined from its sampling distribution; e.g., if $\Pr(t > t_0) = 0.05$, t_0 is the *critical value* of t at the 5 percent level.

curie (Ci): The traditional unit of *radioactivity*. One *curie* (Ci) is equal to 37 billion disintegrations per second (3.7×10^{10} dps = 3.7×10^{10} *Bq*), which is approximately equal to the *decay* rate of one gram of ^{226}Ra. Fractions of a *curie* (e.g. picocurie [pCi], or 10^{-12} *Ci*, and microcurie [µCi], or 10^{-6} *Ci*) are levels typically encountered in *remediation*.

Data Quality Objectives (DQOs): Qualitative and quantitative statements derived from the *Data Quality Objectives (DQO) process* that clarify study technical and *quality* objectives, define the appropriate type of data, and specify tolerable levels of potential decision errors that will be used as the basis for establishing the *quality* and quantity of data needed to support decisions.

delta: (1) As δ, the amount that the distribution of *measurements* for a *survey unit* is increased compared to the distribution of *measurements* of the *reference area*. (2) As Δ, the width of the *gray region*. *Delta* (Δ) divided by sigma (σ), the *arithmetic standard deviation* of the *measurements*, is the *relative shift* expressed in multiples of standard deviations. See in this glossary *relative shift, gray region*.

derived concentration guideline level for small areas of elevated activity ($DCGL_{EMC}$): Based on pathway modeling, the concentration of residual radioactive material within an area of the *survey unit* with elevated activity that corresponds to the *release criteria* (e.g., regulatory limit in terms of *dose* or risk).

derived concentration guideline level for average concentrations over a wide area ($DCGL_W$): Based on pathway modeling, the uniform concentration of residual radioactive material across a *survey unit* that corresponds to the *release criteria* (e.g., regulatory limit in terms of *dose* or risk). This is also known as the wide-area derived concentration guideline level.

detection capability: The *net response level* that can be expected to be seen using a detector with a fixed level of confidence.

detection limit (L_D): The net response level that can be expected to be seen with a detector with a fixed level of confidence.

discrete radioactive particle: Small, usually microscopic, highly radioactive particles having relatively high specific *activity*.

discrimination limit (DL): The level of *radioactivity* selected by the members of the *planning team* that can be reliably distinguished from the *action level*. The *upper bound of the gray region (UBGR)* for *Scenario B* is an example of a *discrimination limit*. See also in this glossary *gray region*, *Scenario A*, and *Scenario B*.

exposure rate: The amount of ionization produced per unit time in air by X-rays or *gamma (γ) radiation*. The unit of *exposure rate* is *roentgens*/hour (R/h); for decommissioning activities, the typical units are microroentgens per hour (µR/h) (i.e., 10^{-6} R/h).

final status survey (FSS): *Measurements* and sampling to describe the radiological conditions of a *site*, following completion of *remediation* activities (if any) in preparation for release. The FSS is the survey in the Radiation Survey and Site Investigation process that is used to demonstrate compliance with release criteria.

gamma (γ) radiation: Penetrating, high-energy, short-wavelength electromagnetic radiation (similar to X-rays) emitted during *radioactive decay*. *Gamma radiation* is very penetrating and requires dense materials (such as lead or steel) for shielding.

gray region: A range of values of the parameter of interest for a *survey unit* where the consequences of making a decision error are relatively minor. In *Scenario A*, the upper bound of the *gray region* is set equal to the *derived concentration guideline level* ($DCGL_W$), and the *lower bound of the gray region (LBGR)* is chosen on a site-specific. In *Scenario B*, the upper bound of the *gray region (UBGR)* is set equal to the *discrimination level*, and the *LBGR* is set equal to the $DCGL_W$.

half-life ($t_{1/2}$): The time in which one half of the atoms of a particular radioactive substance disintegrate into another nuclear form. Also called physical or radiological half-life.

hypothesis: An assumption about a property or characteristic of a set of data under study. The goal of *statistical inference* is to decide which of two complementary *hypotheses* is likely to be true. The *null hypothesis* (H_0) describes what is assumed to be the true state of nature, and the *alternative hypothesis* (H_1) describes the opposite situation.

investigation level: A derived media-specific, *radionuclide*-specific *concentration* that is based on the *release criteria*, that, if exceeded, triggers a response, such as further investigation or *remediation*. See also in this glossary *action level*.

ionizing radiation: High-energy radiation, such as a stream of x-rays, capable of ionizing the substances through which it passes.

lower bound of the gray region (LBGR): The *radionuclide concentration* or level of *radioactivity* that corresponds with the lowest value in the range where the consequence of decision errors is relatively minor. For *Scenario A*, the *LBGR* corresponds is chosen to represent a conservative estimate of the concentration of residual radioactive material. For *Scenario B*, the *LBGR* corresponds to the *derived concentration guideline level (DCGL$_W$)*.

lower limit of detection (LD): The smallest *concentration* of radioactive material in a *measurement* that will yield a net count (above background) that will be detected with at least 95 percent probability and with no greater than a 5 percent probability of falsely concluding that a background observation represents a real signal.

m: (1) As used to describe *measurement* processes, the number of *measurements* from the *reference area* used to conduct a statistical test. (2) As used for a unit of measurement, meters.

measurement method uncertainty: See in this glossary *method uncertainty (u_M)*.

Measurement Quality Objectives (MQOs): *Measurement Quality Objectives (MQOs)* are the specific analytical data requirements of the *Data Quality Objectives (DQOs)*.

method specificity: The ability of the method to measure the radionuclide of concern in the presence of interferences.

method uncertainty (u_M): The predicted uncertainty of the measured value that would be calculated if the method were applied to a hypothetical *sample* with a specified *concentration*.

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

minimum detectable count rate (MDCR): The a priori count rate that a specific instrument and technique can be expected to detect.

nonparametric 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 (WRS) test* and the *Sign test* are examples of nonparametric tests.

non-statistical uncertainties: also known as Type B uncertainties.

NORM: Naturally occurring radioactive material, such as materials containing any of the *radionuclides* produced during the formation of the earth or by interactions of terrestrial matter with cosmic rays as they occur in nature. Examples include radium, uranium, thorium, potassium, and their radioactive *decay products* that are undisturbed as a result of human activities.

normal (gaussian) distribution: A family of bell-shaped distributions described by the *mean* and variance.

power $(1-\beta)$: The probability of rejecting the *null hypothesis* when it is false. The power is equal to one minus the *Type II decision error* rate (i.e., $(1-\beta)$).

power curve: A graph of the *power* as a function of the true value of the parameter of interest. See also in this glossary *power.*

precision: One of the historical data quality indicators (DQIs) recommended for quantifying the amount of error in survey data. Precision represents that portion of the measurement method uncertainty due to random uncertainty.

Quality Assurance Project Plan (QAPP): A written document outlining the procedures a monitoring project will use to ensure the data it collects and analyzes meets project requirements.

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

quality indicators: 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.

Quality Management Plan (QMP): A formal document that describes the *quality* system in terms of the organizational structure, functional responsibilities of management and staff, lines of authority, and required interfaces for those planning, implementing, and assessing all activities conducted.

quality system: A structured and documented management system describing the policies, objectives, principles, organizational authority, responsibilities, accountability, and implementation plan of an *organization* for ensuring *quality* in its work *processes*, products (items), and services. The quality system provides the framework for planning, implementing, and assessing work performed by the *organization* and for carrying out required *quality assurance (QA)* and *quality control (QC)*.

quantile test: A statistical test used in *Scenario B* to identify *areas* of non-uniform contamination.

radioactivity: The property possessed by some elements (such as uranium) of spontaneously emitting energy in the form of radiation as a result of the *decay* (or disintegration) of an unstable atom. Also the *mean* number of nuclear transformations occurring in a given quantity of radioactive material per unit time. The International System (SI) unit of *radioactivity* is the *becquerel (Bq)*. The traditional unit is the *curie (Ci)*.

radiological survey: *Measurements* of radiation levels and *concentrations* of radioactive material associated with a *site* together with appropriate documentation and data evaluation.

ranked set sampling: A two-phase statistical sampling technique in which a subset of statistical *samples* is selected from a larger set of samples based on the rank of the *samples* with respect to the parameter of interest based on professional judgment or, in the case of MARSSIM, some type of field measurement.

relative shift (Δ/σ): *Delta* (Δ) divided by sigma (σ), the *standard deviation* of the *measurements*. See in this glossary *delta*.

relative standard deviation: See in this glossary *coefficient of variation*.

release criteria: Regulatory limits that a *survey unit* must meet before it can be released, expressed either in terms of the *dose* or risk to a future occupant of the *site* or as concentration of radioactive material specified by the applicable *regulation* or standard.

rem (roentgen equivalent man): The traditional unit of *dose equivalent*. The corresponding International System (SI) unit is the *sievert (Sv)*: 1 Sv = 100 rem.

remedial action: An action consistent with a permanent *remedy* either instead of or in addition to a *removal* action in the event of a release or threatened release of a hazardous substance into the environment. A *remedial action* is intended to prevent or minimize the release of hazardous substances so that they do not migrate and cause substantial danger to present or future public health or welfare or the environment.

remediation: Cleanup or other methods used to remove or contain hazardous materials. *Remediation* includes those actions that are consistent with a permanent *remedy* instead of or in addition to a *removal* action in the event of a release or threatened release of a hazardous substance into the environment. *Remediation* is intended to prevent or minimize the release of hazardous substances so that they do not migrate to cause substantial danger to present or future public health or welfare or the environment.

remediation control survey: A type of *survey* that includes monitoring the progress of *remedial action* by real time *measurement* of *areas* being remediated to determine whether efforts are effective and to guide further *remediation* activities.

Scenario A: Scenario that uses a *null hypothesis* that assumes the *concentration* of radioactive material in the *survey unit* exceeds the *derived concentration guideline level* ($DCGL_W$). *Scenario A* is sometimes referred to as "presumed not to comply" or "presumed not clean."

Scenario B: Scenario that uses a *null hypothesis* that assumes the level of *concentration* of radioactive material in the *survey unit* is less than or equal to the discrimination level. *Scenario B* is sometimes referred to as "*indistinguishable from background*" or "presumed clean."

scoping survey: A type of *survey* that is conducted to identify: (1) *radionuclides* present, (2) relative *radionuclide* ratios, and (3) general *concentrations* and extent of *residual radioactive material*.

shift: See in this glossary *delta (Δ)*.

sievert (Sv): The special name for the International System (SI) unit of *dose equivalent*. 1 Sv = 100 *rem* = 1 joule per kilogram (J/kg).

Sign test: A *nonparametric* statistical test used to demonstrate compliance with the *release criteria* when the *radionuclide* of interest is not present in background. See also in this glossary *Wilcoxon Rank Sum (WRS) test*.

simple random sampling: A sampling technique where the *samples* are selected from a larger population in which each *sample* is chosen entirely by chance and each member of the population (i.e., *sample* or *measurement* location) has an equal chance of being selected.

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.

site reconnaissance: A visit to the *site* to gather sufficient information to support a *site* decision regarding the need for further action or to verify existing *site* data. *Site reconnaissance* is not a study of the full extent of *residual radioactive material* at a facility or site or a risk assessment.

Type A: A method of evaluation of *uncertainty* by the statistical analysis of a series of observations. An *uncertainty* component obtained by a *Type A* evaluation is represented by a statistically estimated *standard deviation*, where the standard *uncertainty* is equal to the *standard deviation*.

Type B: A method of evaluation of *uncertainty* by means other than the statistical analysis of series of observations. An *uncertainty* component obtained by a *Type B* evaluation is represented by a quantity that may be considered an approximation to the corresponding *standard deviation*.

Type I decision error: A decision error that occurs when the *null hypothesis* is rejected when it is true. The probability of making a *Type I decision error* is represented by *alpha* (α).

Type II decision error: A decision error that occurs when the *null hypothesis* is accepted when it is false. The probability of making a *Type II decision error* is represented by *beta* (β).

Unity Rule (mixture rule): A rule applied when more than one *radionuclide* is present at a *concentration* that is distinguishable from background and where a single *concentration* comparison does not apply. In this case, the mixture of *radionuclides* is compared against default *concentrations* by applying the Unity Rule. This is accomplished by determining: (1) the ratio between the *concentration* of each *radionuclide* in the mixture, and (2) the *concentration* for that *radionuclide* in an appropriate listing of default values. The sum of the ratios for all *radionuclides* in the mixture should not exceed 1.

upper bound of the gray region (UBGR): The *radionuclide concentration* or level of *radioactivity* that corresponds with the highest value in the range where the consequence of decision errors is relatively minor. For *Scenario A*, the *UBGR* is set equal to the *derived concentration guideline level (DCGLW)*. For *Scenario B*, the *UBGR* is set equal to the *discrimination level*.

W_r: The sum of the ranks of the adjusted *measurements* from the *reference area*, used as the *test statistic* for the *Wilcoxon Rank Sum (WRS) test*.

W_s: The sum of the ranks of the *measurements* from the *survey unit*, used with the *Wilcoxon Rank Sum (WRS) test*.

weighting factor (WT): Multiplier of the equivalent dose to an organ or tissue used for radiation protection purposes to account for different sensitivities of different organs and tissues to the induction of stochastic effects of radiation.

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

REFERENCES

Abelquist, E. 2008. "Dose Modeling and Statistical Assessment of Hot Spots for Decommissioning Applications," Ph.D. Dissertation, University of Tennessee, Knoxville, TN.

Abelquist, E. 2010. *Decommissioning Health Physics, A Handbook for MARSSIM Users*, Taylor & Francis Group, New York, NY.

Abelquist, E. 2014. *Decommissioning Health Physics, A Handbook for MARSSIM Users, Second Edition*, Taylor and Francis Group, Boca Raton, FL.

ANSI/AARST (American National Standards Institute/ American Association Radon Scientists Technologists). 2014. *Protocol for Conducting Measurements of Radon and Radon Decay Products in Schools and Large Buildings*, ANSI/AARST MALB-2014, 34 p. https://standards.aarst.org/MALB-2014/index.html

ANSI/AARST (American National Standards Institute/ American Association Radon Scientists Technologists). 2015. *Performance Specifications for Instrumentation Systems Designed to Measure Radon Gas in Air*, ANSI/AARST MS-PC-2015, 24 p. https://standards.aarst.org/MS-PC-2015/index.html

ANSI/AARST (American National Standards Institute/ American Association Radon Scientists Technologists). 2019. *Radon Measurement Systems Quality Assurance*, ANSI/AARST MS-QA-2019, 24 p. https://standards.aarst.org/MS-QA-2019/index.html

ANSI/AARST (American National Standards Institute/ American Association Radon Scientists Technologists). (2020) Protocol for the Collection, Transfer and Measurement of Radon in Water, ANSI/AARST MW-RN-2020, 48 p. https://standards.aarst.org/MW-RN-2020/index.html

ANSI/IEEE (American National Standards Institute/Institute of Electrical and Electronics Engineers) 2021. N42 Series of American National Standards (various publications and dates). Available at https://webstore.ansi.org/. Accessed on March 3, 2021.

ANSI (American National Standards Institute) 1995. *Traceability of Radioactive Sources to the National Institute of Standards and Technology (NIST) and Associated Instrument Quality Control*. ANSI N42.22-1995. Available at https://webstore.ansi.org/. Accessed on March 9, 2021.

Azami, K., T. Ootagaki, M. Ishida, Y. Sanada. 2018. Characteristics of radiocesium contamination of dy riverbeds due to the Fukushima Daiichi Nuclear Power Plant accident assessed by airborne radiation monitoring. *Landscape and Ecological Engineering*, Vol. 14, pp. 3-15.

Baskaran, M. 2016. Radon Measurement Techniques, *in*: Baskaran, M. (Ed.), *Radon: A tracer of geological, geophysical and geochemical studies*. Springer International Publishing, Switzerland, pp. 15-35. https://standards.aarst.org/MW-RN-2020/index.html

Brodsky, A.1992. "Exact Calculation of Probabilities of False Positives and False Negatives for Low Background Counting," *Health Physics*, 63(2):198-204.

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

Currie, L.A. 1984. Lower limit of detection: Definition and elaboration of a proposed position for radiological effluent and environmental measurements. Washington, DC. U.S. Nuclear Regulatory Commission; Report No. NUreg.cr 4007.

Falciglia, P.P., L. Biondi, R., Catalano, G., Immè, S., Romano, F.G.A. Vagliasindi. 2018. Preliminary investigation for a quasi-quantitative characterization of soils contaminated with 241Am and 152Eu by low-altitude unmanned aerial vehicles (UAVs) equipped with small size γ-ray spectrometer: detection efficiency and minimum detectable activity (MDA) concentration assessment. *Journal of Soils and Sediments*, Vol. 18, pp. 2399-2409.

Falkner, J.T., C.M. Marianno. 2021. Validating a methodology that associates minimum detectable activity with detector velocity. *Health Physics* published ahead of print on March 6, 2021. DOI: 10.1097/HP.0000000000001406.

ICRP (International Commission on Radiological Protection). 2012. *ICRP statement on tissue reactions and early and late effects of radiation in normal tissues and organs – threshold doses for tissue reactions in a radiation protection context.* ICRP Publication 118. Ann. ICRP 41(1/2).

IEC (International Electrotechnical Commission). 2000. *Radiation Protection Instrumentation - Radon and Radon Decay Product Measuring Instruments - Part 2: Specific Requirements for Radon Measuring Instruments*, IEC 61577-2, International Electrotechnical Commission, Geneva, Switzerland. https://standards.globalspec.com/std/1694803/IEC%2061577-2

IEC (International Electrotechnical Commission). 2006. *Radiation Protection Instrumentation - Radon and Radon Decay Product measuring Instruments - Part 1: General principles*, IEC 61577-1, International Electrotechnical Commission, Geneva, Switzerland. https://standards.globalspec.com/std/380787/IEC%2061577-1

IEC (International Electrotechnical Commission). 2014. *Radiation protection instrumentation - Radon and radon decay product measuring instruments - Part 3: Specific requirements for radon decay product measuring instruments*, IEC 61577-3, International Electrotechnical Commission, Geneva, Switzerland. https://standards.globalspec.com/std/9897537/ds-en-61577-3

ISO (International Organization for Standardization) 2012. *International Vocabulary of Basic and General Terms in Metrology* (VIM), third edition.

ISO (International Organization for Standardization) 2015a. *Reference materials – Selected terms and definitions.* ISO Guide 30:2015. Geneva, Switzerland.

ISO (International Organization for Standardization) 2015b. *Uncertainty of measurement — Part 1: Introduction to the expression of uncertainty in measurement, ISO Guide to the Expression of Uncertainty in Measurement (GUM)* ISO/IEC GUIDE 98-1:2009

ISO (International Organization for Standardization). 2019. ISO 11665-1:2019 - *Measurement of radioactivity in the environment -- Air: radon-222*, p. 33. https://standards.iteh.ai/catalog/standards/iso/950298e5-4976-418a-9b5b-d8d4aa7ad47e/iso-11665-1-2019

Ji, Y.Y., T. Lim, K. Hitomi, T. Yajima. 2020. Spectroscopic Estimation of Dose Rate Induced from Radioactive Cesium in the Ground Using a Mobile Gamma-Ray Spectrometry Based on a $LaBr_3$(Ce) Detector. *Health Physics*, Vol. 118, No. 2, pp. 215-225.

Ji, Y.Y., T. Lim, H.-Y. Choi, K.H. Chung, M.J. Kang. 2019. Development and Performance of a Multipurpose System for the Environmental Radiation Survey Based on a $LaBr_3$(Ce) Detector. *IEEE Transactions on Nuclear Science*, Vol. 66, No. 12. Pp 2422-2429.

Kock, P., C. Rääf, C. Samuelsson. 2014.On background radiation gradients – the use of airborne surveys when searching for orphan sources using mobile gamma-ray spectrometry. *Journal of Environmental Radioactivity*. Vol. 128, pp. 84-90.

Kock, P., C. Samuelsson. 2011. Comparison on airborne and terrestrial gamma spectroscopy measurements – evaluation of three areas in southern Sweden. *Journal of Environmental Radioactivity*, Vol. 102, pp. 605-613.

Lee, C., H.R. Kim. 2019. Optimizing UAV-based radiation sensor systems for aerial surveys. *Journal of Environmental Radioactivity*, Vol. 204, pp. 76-85.

Marianno, C.M. 2015. Signal processing and its effect on scanning efficiencies for a field instrument for detecting low-energy radiation. *Health Physics*, Vol. 109, pp. 78-83.

Marques, L., A. Vale, P. Vaz. 2021. State-of-the-Art Mobile Detection Systems for Different Scenarios. *Sensors*, Vol. 21, No. 1051, p. 1051.

NCRP (National Council on Radiation Protection and Measurements) 1989. *Limit for exposure to "hot particles" on the skin.* NCRP Report No. 106.

NIST (National Institute of Standards and Technology). 1994. *Guidelines for Evaluating and Expressing Uncertainty of NIST Measurement Results.* (NIST Technical Note 1297). Gaithersburg, MD. National Institute of Standards and Technology.

NIST (National Institute of Standards and Technology). 2020. Metrological Tools for the Reference Materials and Reference Instruments of the NIST Material Measurement Laboratory. NIST Special Publication 260-136. Gaithersburg, MD.

NJDEP (New Jersey Department of Environmental Protection). 2005. Chapter 12, Radiological Assessment, in *Field Sampling Procedures Manual*, Trenton, NJ, 18 p. (available online https://www.nj.gov/dep/srp/guidance/fspm/pdf/chapter12.pdf).

NRC (U.S. Nuclear Regulatory Commission). 1987. *Control of hot particle contamination at nuclear power plants.* US Nuclear Regulatory Commission. Information Notice 87-39.

NRC (U.S. Nuclear Regulatory Commission). 1998.. *A Nonparametric Statistical Methodology for the Design and Analysis of Final Status* 6 *Decommissioning Surveys.* NUREG-5 1505

NRC (Nuclear Regulatory Commission) 2020. *Minimum Detectable Concentrations with Typical Radiation Survey for Instruments for Various Contaminants and Field Conditions.* NUREG-1507, Revision 1.

Peeva, A. 2021. Now Available: *New Drone Technology for Radiological Monitoring in Emergency Situations.* Published February 1, 2021. International Atomic Energy Agency (IAEA). IAEA Department of Nuclear Sciences and Applications. Vienna, Austria. Available at www.iaea.org. Accessed on March 3, 2021.

Rahman, N.A.A., K.S.M. Sahari, M.F.A. Jalal, A.A. Rahman, M.I.A. Adziz, M.Z. Hassan. 2020. Mobile robot for radiation mapping in indoor environment. *IOP Conference Series: Materials Science and Engineering*, Vol. 785, 012021.

Sanada, Y., T. Torii. 2015. Aerial radiation monitoring around the Fukushima Daiichi nuclear power plant using an unmanned helicopter. *Journal of Environmental Radioactivity*, Vol. 139, pp. 294-299.

Sanada, Y., Y. Urabe, M. Sasaki, K. Ochi, T. Torii. 2019. Evaluation of ecological half-life of dose rate based on airborne radiation monitoring following the Fukushima Dai-ichi nuclear power plant accident. *Journal of Environmental Radioactivity*, Vol 210, 105816.

Sanderson, D. 2013. Measuring regional scale distribution of radiocaesium. Presented at the Caesium workshop: "Fukushima recovery–understanding, modelling and managing radiocaesium decontamination, COEASSE," Fukushima, Japan. September 30–October 3, 2013. Available at https://fukushima.jaea.go.jp/fukushima/result/pdf/pdf00/24__Sanderson.pdf. Accessed on March 3, 2021.

Sinclair, L.E., R. Fortin, J.L. Buckle, M.J. Coyle, R.A. Van Brabant, B.J.A. Harvey, H.C.J. Seyward, M.W. McCurdy. 2016. Aerial Mobile Radiation Survey Following Detonation of a Radiological Dispersal Device. *Health Physics*, Vol. 110, No. 5, pp. 458-470.

Subba, Ramu MC, Raghavayya M, Paul AC. 1994. *Methods for the measurement of radon, thoron and their progeny in dwellings*, AERB Technical Manual, TM/RM – 1

Tanigaki, M., R. Okumura, K. Takamiya, N. Sato, H. Yoshino, H. Yamana. 2013. Development of a car-borne γ-ray survey system, KURAMA. *Nuclear Instruments and Methods in Physics Research* A, Vol. 726, pp. 162-168.

U.S. EPA (Environmental Protection Agency) 1980. *Upgrading Environmental Radiation Data*. Health Physics Society Committee Report HPSR-1, Office of Radiation Protection, Washington, DC

U.S. EPA (Environmental Protection Agency) 1994. *Statistical Methods for Evaluating the Attainment of Cleanup Standards, Volume 3: Reference Based Standards for Soils and Solid Media.* Office of Policy, Planning, and Evaluation, EPA, Washington, DC (EPA 230-R-94-004, PB94-176831).

U.S. EPA. (Environmental Protection Agency) 2000. *Multi-Agency Radiation Survey and Site Investigation Manual, Revision 1*. MARSSIM, Washington, DC NUREG-1575, Rev. 1, EPA 402-R-97-016, Rev. 1 DOE/EH-0624, Rev. 1.

U.S. EPA. (Environmental Protection Agency) 2004. MARLAP. *Multi-Agency Radiological Laboratory Analytical Protocols Manual*. Volumes 1 - 3. Washington, DC: EPA 402-B-04-001A-C, NUREG 1576, NTIS PB2004-105421. (July 2004). Available at: https://www.epa.gov/radiation/multi-agency-radiological-laboratory-analytical-protocols-manual-marlap

U.S. EPA. (Environmental Protection Agency) 2020. *Multi-Agency Radiation Survey and Site Investigation Manual, Revision 2*. (Draft for Public Comments) (MARSSIM), Washington, DC, EPA-402-P-20-001.

Vaupotič, Janja, Smrekar, Nataša, Žunić, Z.S. 2017. Comparison of radon doses based on different radon monitoring approaches: *Journal of Environmental Radioactivity*, V. 169–170, P. 19-26.
https://doi.org/10.1016/j.jenvrad.2016.11.023

Vyletělová, Petra, Froňka, Aleš. 2019 Continuous radon-in-water monitoring—comparison of methods under laboratory conditions and results of in situ measurements: Radiation Protection Dosimetry, V. 186(2-3), P. 406–412. https://doi.org/10.1093/rpd/ncz241

WHO (World Health Organization) 2009. Radon measurements: *in* editors, Hajo Zeeb, and Ferid Shannoun, WHO handbook on indoor radon: a public health perspective, Geneva, Switzerland, p. 21-40.

Wilhelm, E., N. Arbor, S. Gutierrez, S. Ménard, A.-M. Nourreddine. 2017. A method for determining Am-241 activity for large area contamination. *Applied Radiation and Isotopes*, Vol. 119, pp. 86-93.

APPENDIX A: EPA'S CHARGE QUESTIONS

EPA SCIENTIFIC ADVISORY BOARD – RADIATION ADVISORY COMMITTEE CHARGE TO THE PANEL – MARSSIM, REVISION 2

The Multi-Agency Radiation Survey and Site Investigation Manual (MARSSIM) document (https://www.epa.gov/radiation/multi-agency-radiation-survey-and-site-investigation-manual-marssim) provides information on planning, conducting, evaluating and documenting building surface and surface soil[1] final status radiological surveys. MARSSIM is a multi-agency consensus document that was developed collaboratively by four Federal agencies having authority and control over radioactive materials: Department of Defense (DOD), Department of Energy (DOE), Environmental Protection Agency (EPA) and Nuclear Regulatory Commission (NRC). The MARSSIM document's objective is to describe a consistent approach for planning, performing and assessing building surface and surface soil final status surveys to meet established dose- or risk- based release criteria, while at the same time encouraging an effective use of resources.

The original MARSSIM document was published in 1997, with errata and addenda pages published in 1998 and 1999. Revision 1 to MARSSIM was published in 2000, and additional errataand addenda pages were published in 2001. None of the changes made from 1998 to 2001 reflect significant departures from the science and technology of the original MARSSIM document; instead, they provide additional clarification and correct errors in the original published document. No additional changes to the document itself have been made since 2001.

The Scientific Advisory Board (SAB) conducted the scientific peer reviews of the 1997 version of MARSSIM (EPA-SAB-RAC-97-008, dated 9/30/1997), its companion document addressing laboratory analytical protocols (MARLAP, EPA-SAB-RAC-03-009, dated 6/10/2003) and the MARSSIM Supplement addressing materials and equipment (MARSAME, EPA-SAB-08-010, dated 8/7/2008).

The MARSSIM Workgroup developed a three-day in-person or five-day (4 hours per day) internet-based technical training course on the document for radiation professionals seeking to learn more about final status surveys for surface soils and building surfaces. The EPA-sponsored MARSSIM training is offered three times a year to a total of 72 students.

The MARSSIM Workgroup conducted a thorough request for public input for the MARSSIM revisions in 2010. In addition, the MARSSIM Workgroup held a Consultation with the SAB in 2011 to request input on possible updates. After developing a draft of the proposed document, the MARSSIM Workgroup conducted an Internal Agency Review in 2016, which identified further areas of clarification and improvement. Finally, the MARSSIM Workgroup plans to make Revision2 available for public comment and will incorporate suggested improvements as appropriate based on that review.

Previous scientific peer reviews have helped to shape the science behind the MAR- series of documents, and the four federal agencies involved in the MARSSIM Workgroup agree that input from the SAB should be sought for any significant changes, including those currently proposed forRevision 2 (outlined in the charge questions below). Scientific concepts remaining materially unchanged from Revision 1, (e.g., the use of non-parametric statistics, Scenario A) have already undergone review by the SAB and do not require review at this time.

[1] The MARSSIM document does not address volumetric or subsurface soils.

CHARGE QUESTIONS:

1) Are the revisions to MARSSIM concepts and methodologies technically appropriate, usefuland clear, and do they provide a practical and implementable approach to performing environmental radiological surveys of surface soil and building surfaces?

1.1 Please identify whether the inclusion and proposed implementation of scan-only surveys (Section 5.3.6.1 and Section 8.5) is appropriate, adequate and clear, especially the discussion onsampling for scan-only measurement method validation or verification.

> The MARSSIM Workgroup wrote MARSSIM, Revision 1, for 1995 technology, not envisioning that future instrumentation would be able to measure a statistically significantportion of the survey unit while meeting required Measurement Quality Objectives (MQOs), especially that the Minimum Detectable Concentration (MDC)/Minimum Detectable Activity (MDA) be less than 50% of the Derived Concentration Guidelines Level for wide areas (DCGLw). New methods for designing, implementing and assessingscan-only surveys are included in the revisions to make effective use of resources when employing these technologies.
>
> Earlier reviewers misinterpreted the term "scan-only surveys" to mean that samples wouldnot be taken as any part of the survey process. Revision 2 has been further revised to indicate that quality control samples may need to be collected as part of the method validation or verification process, as appropriate.

1.2 Please comment on the inclusion and proposed implementation of Scenario B (Chapter 4, Section 5.3, and Chapter 8). Is it appropriate to recommend that Scenario B be used only for those situations where Scenario A is not feasible? Are methods for considering background variability inassessing whether the site is indistinguishable from background reasonable and technically accurate? Is the inclusion and proposed implementation of added requirements for retrospective power analysis and the Quantile Test while using Scenario B technically appropriate and discussedadequately and clearly?

> Under hypothesis testing in MARSSIM, Scenario B is defined as assuming that the survey unit meets the release criteria unless proven otherwise, and its use was discouraged in MARSSIM, Revision 1. However, this is the only viable option for sites where the criterionis effectively "no added radioactivity" or "indistinguishable from background".
>
> In Scenario B, the Lower Bound of the Gray Region (LBGR) is often set to zero, but thedocument allows use of a non-zero LBGR that considers background variability in determining whether the survey unit is indistinguishable from background.

Since Scenario B assumes that the site meets the release criteria, there is a risk that the survey unit will pass simply because the survey did not have sufficient rigor. To guard against that, the revisions require that when using Scenario B, the survey unit must per-forma retrospective power analysis to prove the survey has sufficient statistical power to detect asurvey unit that should not have passed.

The non-parametric tests included in MARSSIM test the median instead of the mean. The release criteria are typically expressed as the mean. To guard against Scenario B situations where the median will pass but the mean won't (this can occur in sample data distributions with a long tail in the higher concentration range), Revision 2 also requires that when usingScenario B, the survey unit must pass a quantile test to guard against excessive skewness.

1.3 Is the proposed implementation of the of the concept of Measurement Quality Objectives adequately and correctly described, including the concept of measurement uncertainty (Chapter 4 and Appendix D)? Is the proposed calculation of measurement uncertainty consistent with theconcept of Measurement Quality Objectives? Is the method appropriate and practical for both laboratory and field (including scan) measurements? Please comment on the concerns of stakeholders that calculating measurement uncertainty for field measurements makes the survey process difficult to implement. In addition, please comment on whether recommendations providedby NIST, ANSI/IEEE and MARLAP for measurement quantifiability should be incorporated further into MARSSIM, Revision 2, or whether the current recommendations should be left as is (e.g., the original MARSSIM requirement that the MDC/MDA should be set at 10-50% of the action level).

The concept of MQOs as a subset of Data Quality Objectives (DQOs) originated after publication of MARSSIM, Revision 1. The use of MQOs ensures that each measurement taken is of sufficient quality to be used as part of the survey design. These MQOs include many familiar Data Quality Indicators, which were included in MARSSIM, Revision 1, such as range, specificity, ruggedness and detection capability, typically represented as MDC/MDA. However, the older Data Quality Indicators of bias and precision have been captured by a new MQO: measurement uncertainty, with bias indicating systematic uncertainty and precision indicating random uncertainty. The International Organization forStandardization published the Guide to Uncertainty in Measurement in 1995. The National Institute for Standards and Technology (NIST) published Technical Note 1297: Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results (GUM) in 1994, which provided guidance to the federal government to incorporate measurement uncertainty into their procedures. As a result, subsequent MAR-series documents MARLAPand MARSAME included information on the use of measurement uncertainty.

MARSSIM, Revision 1, indicated that the greater source of error for a survey was typicallyfound in the sampling design, not in the measurements themselves, and as a result, did not emphasize concerns regarding measurement uncertainty. However, with the inclusion of scan-only surveys, the sampling design error decreases significantly as a greater percentageof the survey unit is covered. Consequently, the measurement error becomes critical, and thus the more quantitative method of assessing and controlling measurement uncertainty similarly becomes critical. Stakeholders have expressed concerns that calculating measurement uncertainty, specifically for field measurements, makes the survey process difficult to implement. The MARSSIM Workgroup agreed to include the MQO for measurement uncertainty and investigate future tools to make process easier.

The American National Standards Institute/Institute of Electrical and Electronics Engineers(ANSI/IEEE) standard N42.23 recommends that interpretation of survey data involving environmental media, "such as soil, sediments, concrete and water, should not use the MDC/MDA to evaluate measurement results, and instead recommends use of the decision level or considering the confidence interval for the measurement result." The authors of MARSSIM, Revision 1, understood that for cases when the decision to be made concerns the mean of a population that is represented by multiple measurements, detection criteria based on the MDC/MDA may not be sufficient and a somewhat more stringent requirementwas needed. To meet this need, they introduced an additional requirement that the MDC/MDA should be set at 10-50% of the action level. This predated the concept of measurement quantifiability (as considered in MARLAP and ANSI/IEEE N42.23), but it results in comparable constraints on a Minimum Quantification Concentration (See MARSAME Section 7.6 for further discussion.) To minimize changes to current practice, the original MARSSIM requirement is left as is in Revision 2.

1.4. Is the discussion of survey requirements for areas of elevated activity technically accurate,appropriate and clear? In particular, please comment on the decision to maintain the use of the unity rule for multiple areas of elevated activity (Section 5.3.5, Section 8.6 and Appendix O.4).Are there suggested alternatives to the use of the unity rule?

While modeling is outside the scope of MARSSIM, depending on the modeling tool or methodology used to develop release criteria, the use of the Unity Rule for multiple areas ofelevated activity in a single survey unit can lead to unrealistic or overly conservative assumptions. For example, the models may assume that the receptor is located directly above each area of elevated activity and stays there for the duration of their exposure period. This physically cannot occur in cases where there is more than one area of elevated activity per survey unit and results in concerns that this will cause an over-estimate of dose or risk, leading to an emphasis on remediating areas of elevated activity that don't incur additional significant dose or risk to receptors.

MARSSIM, Revision 2, does not change recommendations for the use of the unity rule, but emphasizes assessing whether criteria for areas of elevated activity apply to survey units, and when they do, using a commonsense approach to applying these criteria, keeping in mind the limitations of the unity rule described above for multiple areas of elevated activity.

1.5. Is the discussion of the use of MARSSIM surveys for addressing sites containing discrete radioactive particles technically sound and appropriate, and is the description accurate?

In particular, please comment on the rule-of-thumb for determining when use of MARSSIM may notbe appropriate for survey units containing discrete radioactive particles (Section 4.12.8 and Appendix O.5).

> Discrete radioactive particles have an extremely small size and contain enough activity that survey units containing discrete radioactive particles generate impractical survey designs under MARSSIM. Over MARSSIM's twenty-year history, several sites have attempted to utilize MARSSIM to address discrete radioactive particles, with predictably extreme surveydesigns as a result. In addition to being impractical, designs for discrete radioactive particles violate some of the assumptions commonly made during modeling, which includesparameters based on an areal source of radioactive material, e.g., length of the area of the elevated activity in the direction of overland flow. While modeling is outside of the scope of MARSSIM, it is nonetheless required that survey designs match the assumptions made during modeling, otherwise, the survey design does not meet the requirements of the action level.
>
> To set a limit for determining when areas of elevated activity are too small to use the traditional MARSSIM methodology, the MARSSIM Workgroup used a traditional rule-of- thumb for instrumentation. When the length of the area of elevated activity is less than three times the distance to the detector, the area of elevated activity is viewed by the detector as a point source instead of as an areal source. These point sources will need different receptor modeling and release requirements, and hence different survey designs than traditional arealsources.
>
> At this time, MARSSIM does not provide guidance on designing discrete radioactive material surveys. It is the intention of the revision that additional information provided should prevent MARSSIM from being applied inappropriately to survey units involvingdiscrete radioactive particles.

2) Does MARSSIM, Revision 2 provide useful, appropriate and clear examples and descriptions of technical approaches to implementing surveys and the statistics by which theyare interpreted?

2.1 Please comment on whether the description of updated measurement methods and instrumentation information (Chapter 6 and Appendix H) is useful, appropriate and clear.

2.2. Please comment on whether the additional optional methodology for the use of Ranked SetSampling (Appendix E) for hard-to-detect radionuclides is useful, appropriate and clear.

> The Ranked Set Sampling methodology requires a close, reasonable and provable correlation between an easy-to-measure attribute of the sample (e.g., soil sample size distribution) and theactivity level of a hard-to-detect radionuclide. While challenging to implement in practice, therevisions include this optional method to assist sites with designing surveys for hard-to-detect radionuclides, which can be difficult and resource intensive to implement.

2.3 Please comment on whether the new and additional examples provided in Chapter 5 areuseful, appropriate and clear.

3) Is the information in MARSSIM, Revision 2 clear, understandable and presented in a logicalsequence? How can the presentation and content of material be modified to improve the understandability of the manual?

3.1. Please comment on the revised description of how to set the Lower Bound of the GrayRegion (LBGR) and its likely effectiveness in encouraging users to rely on site-specific information for doing so (Chapter 4 and Section 5.3).

> One of the critical decisions made during site survey design under MARSSIM Scenario A is to set a value for the LBGR. Twenty years of training and review of survey plans have shown that this concept is not well understood by users, and that users tend to implement the standard rule of thumb of setting the LBGR to 50% of the DCGLw. This rule of thumb was provided in MARSSIM, Revision 1, for use only when additional information was not available. A poorly chosen value for the LBGR can affect the power of a survey resulting inunnecessary use of resources or a higher chance of failing a survey unit that meets the release criteria.
>
> In Scenario A, the LBGR should be set equal to a conservative estimate of the average concentration remaining in the survey unit. This information is typically available from historical site information, or a scoping or characterization survey if the survey unit is un-remediated, or the remedial action survey if the site has been remediated. The purpose of the revisions is to describe this concept in plain language, moving away from a statistics terminology description of the concept.

3.2. Please comment on whether avoiding the use of the term "area factor" improvesunderstandability of the elevated measurement comparison concept (Section 8.6.1).

> Area factors, which are simply the ratio of the Elevated Measurement Comparison (EMC) release criteria to the wide-area release criteria, should be based on site-specific modeling or calculations. Due to the misapplication of published area factors from the literature and to provide focus on the need for development of site-specific EMC criteria, MARSSIM, Revision 2 avoids the use of the term area factor. In addition, lessons learnedfrom training MARSSIM show that describing the EMC concept in descriptive language,rather than by defining additional terminology, seems to improve understandability of theconcept.

3.3 Please comment on the effectiveness of the new organization of Chapter 4 (Considerations forPlanning Surveys) to improve the understandability of the Chapter.

> Earlier reviews of Chapter 4 provided evidence that the fundamental organization of Chapter 4 made it difficult to find and understand vital information. After discussing

the challenge with experts in training and explaining the material, Chapter 4 was completelyrewritten or reorganized in an attempt to improve understandability without changing thefundamental purpose of or material in the Chapter. In an effort to streamline the presentation of material in Chapter 4, some information was moved to Appendix O.

3.4. Please comment on the effectiveness of moving derivations from Chapter 5 to Appendix O toimprove the understandability of the Chapter.

In an effort to streamline the presentation of material in Chapter 5, some derivations of key concepts were moved to Appendix O.

APPENDIX B: EDITORIAL COMMENTS

Editorial Comments from Response to Charge Question 1.1:

Section 5.3.6: A reference to Table 5.5, which summarizes Section 5.3.6.1, is needed earlier.

Page 5-42, line 25: Recommend quantifying the range of typical scanning coverage to replace the vague statement of "a much larger portion" in comparison to surveys based on discrete sampling and measurement.

Page 5-43, Equation 5-10: It would be more accurate if "Scan Area" were relabeled "% of Scan Area".

Table D.2: When the true condition is exceeds release criterion, Table D.2 for Scenario B describes a decision error from Accepting H_0 as "Incorrectly Fail to Release Survey Unit." However, this decision error would be incorrectly release survey unit.

Table D.2: When the true condition is meets release criterion, Table D.2 for Scenario B describes a decision error from Rejecting H_0 as "Incorrectly Release Survey Unit." However, this decision error would be incorrectly fail to release survey unit.

Page D-25, line 18: Wording that associates Type II errors with a for Scenario B and b for Scenario A doesn't completely agree with the associations shown in Tables D.1 and D.2.

Page D-25, lines 4-5: … site-specific area factors ~~can also~~ **should** be developed.

Editorial Comments from Response to Charge Question 1.2:

Page 5-28, Figure 5-7, Gray Region for Scenario A: The meanings of the abbreviations LBGR and $DCGL_W$ should be added as footnote to the figure.

Page 5-29, Figure 5-8, Gray Region for Scenario B: The meanings of the abbreviations AL and DL should be added as footnote to the figure.

Editorial Comments from Response to Charge Question 1.3:

- The following abbreviations appear nowhere in the text and only in "Symbols, Nomenclature and Notations"
  - y_c
  - y_D

  They should be used in the text or removed from the tables.

- The u symbols [*i.e.,* $u(x_i)$, $u(x_i,x_j)$, $u_c(Y)$, U, c_i] in the Symbols, Nomenclature, and Notations list should be used where appropriate in the text. The u and σ symbols should be used where appropriate, particularly where Δ/u should be used.

- In Figure 4.1, wording inside the information boxes reference sections of Chapter 4 (i.e., 4.3, 4.4, etc.) does not reflect the title or content of those sections of Chapter 4.
- Page 4-4 line 21. Uses the word "authentic" to describe samples. This term is not defined nor is it commonly used in field sampling. Using words like reliable, dependable, trustworthy, or valid are more in keeping with terminology used in such investigations

- Page 4-5 lines 6-8. Information on selecting the number and type of QC measurements for a specific project are provided in Section 3.4: Tables 4, 5, and 6 of the UFP-QAPP Part 1, and Worksheet 28 of the UFP-QAPP Part 2A. Some of these should be included in Appendix D, or even in Chapter 4. This part of the discussion is central to the entire chapter and the whole discussion regarding MQO. This co-location would make it much easier for the user to view the most essential of these tools.

- Page 4-5 line 16. Measurement performance criteria are not defined here. The term is mentioned in Appendix D.2.2, but it is not defined. Definition is needed in Chapter 4.

- Page 4-31 line 37 "MARSSIM recommends that a realistic or conservative estimate of the MDC be used instead of an optimistic estimate." An example of what might be realistic or conservative in the context of uncertainty estimates for MDC would be helpful.

- Page 4-55 line 18. "…alpha of 0.05 and beta of 0.10". Suggest replacing with, "Type-I error (alpha) and Type-II error (beta)".

- Appendix D p. D-53, line 1-3 "The uncertainty of a measurement expressed as combined standard uncertainty includes the counting uncertainty of the measurement instrumentation and the sum of the errors associated with the measurement system." Recommend rewording of the sentence by using the Equation 6-18 terminology that the uncertainties are combined as the root-sum-of-squares.

- Chapter 6, p. 6-31, Example 8 illustrates *some* of the steps used to estimate an uncertainty, in this case σ_y, for a measurement counting process. Recommend detailing all the steps of this process for estimating measurement uncertainties, and also for the theoretical total standard deviation of the population distribution being sampled, σ, that is used for MQO Δ/σ.

- Terminology used throughout MARSSIM should use the guidance of NIST Technical Note 1297 (1994) and ISO (2015b).

- On page 6-2, line 27, “Detection capabilities” is utilized here and then “detection capability” is defined in section 6.3. Page 6-6, line 22 states “The detection capability (sometimes referred to as sensitivity)”. MARSSIM needs to move the definition of ‘capability’ to earlier in the document and use ‘capability’ consistently throughout MARSSIM, including Appendix H

- The *a posteriori* MARSSIM historical measurement records should be used to validate the MDC practice for MARSSIM cleanup projects, i.e., see how applicable and accurate the concept was.

Editorial Comments from Response to Charge Question 1.4:

- Page 4-8, Section 4.3.6 A note on subsurface assessment:
  - Suggest repeating the definition of surface vs subsurface depth, to remind the reader.

- Page 4-9, Section 4.4:
  - The Unity Rule is well described.
  - Line 5. The sentence: Essentially, this means that if measurements….helps clarify the Unity Rule.
  - Line 15-16: to slightly clarify this paragraph, change the sentence to read “…..each fraction, (f), is determined…..”

- Page 4-18, Section 4.5.3.7: Small areas of Elevated activity. Although discussed in Section 5.3.5, (and 4.2.5 as well) elevated activity in multiple areas and the Unity Rule also discussed in this section and is appropriate. It references Sections 5.3.5.1 and 5.3.5.2.

- Page 5-36, line 3 should read “…treatment of areas of elevated radioactive materials…”. The adjective “elevated” is misplaced.

- Page 5-36, line 31, Equations 5-1 and 5-2 – The notation in Section 5.3.5 is inconsistent and confusing. For example, *A* is defined as the “total area of the survey unit”. But in Equations 5-1 and 5-2, the total area of the survey unit now seems to be defined as “*A* (survey unit)”

- Page 5-37, Equations 5-3 and 5-4 – “*A* (grid area)” is used in these equations and never defined. Further “*A* (grid area)” and “*A* (surface area)” are not defined in “Symbols, Nomenclature, and Notations,” Page xxviii. Definitions included on Page xxviii is “*A*” for overall sensitivity of a measurement and “A” is area.

- Page 5-37, line 4 – “A_{EA}” is defined here and is consistent with “Symbols, Nomenclature, and Notations” on Page xxviii as *area of elevated activity*. This as an example of good and logical notation that should be used consistently throughout Section 5.3.2 and the entire MARSSIM document.

- Page 5-37, Equations 5-5 and 5-6 – The use of “Scan MDC (actual)” and “Scan MDC (required)” is confusing in these two equations and never defined. If someone jumps to these equations and starts to apply then without reading the context for which they apply, they will come to the conclusion that Scan MDC (actual) = Scan MDC (required). To avoid the possible confusion and the poor notation, the SAB recommends these two equations be removed. Rather than "Scan MDC (actual)" would it be clearer to the reader if the narrative was "Actual scan MDC" or "Scan MDC (required)"?

- Page 5-41, Example 8, the purpose for the following statement is unclear: "The grid area encompassed by a triangular sampling pattern of 10 m is approximately 86.6 m^2, as calculated using Equation 5-3: The very next line completes the calculation, showing it to be 99.1 m^2. This is confusing, where did the initial estimate of 86.6 m^2 come from?

- Page 8-47, Section 8.6.2, the Unity Rule derivation with regard to $DCGL_{EMC}$ and $DCGL_W$ in Equation 8-4 should be clarified.

- Page 8-45, Section 8.6: Evaluate the Results – The Decision
    - 8.6.3 discusses what to do if the survey unit fails and lays out possible options, with a given example (13). Example 13-15 are appropriately illustrative.
    - Example 15 was not as clear as the others. It uses the Wilcoxon Rank Sum test. Is this the only test available, and are the assumption that go into the WRS appropriate for all Class 1 failures?
    - It would be helpful to mention Appendix O in example 15 noted above.

- Appendix O, page O-6, Section O.4.4 shows tables of example area factors, saying it is strictly for purposes of illustrating the concepts involved, but does not comment on how and why the trends in the tables differ by radionuclide. One assumes this must be due to the ranges of the penetrating radiations emitted. The reader would benefit from a brief explanation on the reasons for table trends and relevant additional information related to it.

- Appendix O, page O-7, line 25. "When applicable, As Low As Reasonably Achievable (ALARA) criteria should be considered". The statement should be expanded to provide an example.

- Appendix O, page O-7, line 28, the Board reaffirms its agreement with the statement in Appendix O stating it is always acceptable and conservative to assume the smallest area factor possible (e.g., AF = 1). Chapter 5 & 8 should use additional call-ins to the appropriate sections in the Appendixes to facilitate the clarification of details in the discussion.

Editorial Comment from Charge Question 1.5:

Section 4.12.8 of Chapter 4 is titled, Release Criteria for Discrete Radioactive Particles. The section offers no release criteria; therefore, the title should be amended to better reflect the contents of the section.

Editorial Comments from Response to Charge Question 2.1:

Editorial comments on Chapter 6:

- Equation 6.2 should read: $L_D = k^2 + 2k\sqrt{2B}$

- Page 6-9, lines 24-26 – To prevent misuse of the MDC as a decision criterion, delete the sentence: "Underestimating an MDC can have adverse consequences, especially if activity is later detected at a level above the stated MDC."

- Page 6-10, Example 1 – Replace "concentration *C* =" with "constant *C* in equation 6-5 is"

- Page 6-11, Table 6.1 – The additional heading "approximate detection capability" should be deleted from the table. Neither L_C or L_D represent an approximate detection capability.

- Page 6-1, line 16 – Punctuation is needed to clarify the double use of "and".

- Page 6-2, line 32 – Recommend wording to be "background of the specific radionuclides of interest". "Specific" should modify radionuclides, not background.

- Page 6-8, above line 7 – Equation 6.2 should read: $L_D = k^2 + 2k\sqrt{2B}$

- Page 6-10, Example 1, line 6 – "concentration" should be replaced with "constant".

- Page 6-10, Example 1 – (1/15 cm^2) should be (1 count/15 cm^2)

- Page 6-35, lines 35, 36 – Use of the descriptor "activator" is not used consistently in Chapter 6 and Appendix H when referring to scintillators, and TLD and OSL materials. One example is being highlighted here, but there are numerous. NaI(Tl) should be referred to as "thallium-activated sodium iodide". The use of " -activated" is not used consistently in these sections of the document.

- Page 6-35, lines 28 - 29 – Consider replacing "neon or helium" with "noble gas" that is more inclusive to the possible gases used. Also consider removing the reference to "quenching agent", just simply state that a small amount of halogen is added. Methane (lines 26-27) is also referred to as a "quench agent", but the two quench agents are quenching different phenomenon. To avoid that level of detail, it would be easiest to remove the reference to the quench agent.

- Page 6-36, line 26 – "sensitivity" should be replaced with "capability".

- Page 6-36, line 1 – Recommend removing "organic" and simply refer them as "plastic scintillators". Or is MARSSIM referring to "plastic scintillators" and "organic scintillators"? Stilbene is an example of an organic scintillator.

- Page 6-42, line 4 – "sensitivity" should be replaced with "capability".

- Page 6-44, line 18-19 – Recommend rewording "…direct measurements and scanning …". Aren't large area detectors primarily (exclusively?) used for scanning?

- Page 6-45, line 18 – "Disposition survey" is used for the second time in MARSSIM here. "Disposition survey" is never defined.

- Page 6-48, Table 6.8 – Recommend spelling out in situ gamma-ray spectrometer rather than using ISGS in the first column.

- Page 6-48, Table 6.8 – "Hand-Held Instrument" "Smear" "Advantages", 1st bullet – Recommend rewording to "Easily transportable measurement technique for assessing removable radioactive material".

- Page 6-49, Table 6.8 – The use of "Laboratory Analysis" as the "Instrument" is not appropriate. Nice to have Laboratory Analyses available for comparison, but would work best as a separate table.
  - "Laboratory Analysis" "Sampling" "Disadvantages", 4th bullet – Although true, also true with hand-held instruments.

- Page 6-50, Figure 6.2 – Figure would be improved if the protective area screen was illustrated in the figure.

- Page 6-50, line 4 – To avoid confusion recommend rewording to "The conversion of instrument display of counts to surface activity in units of Bq/m^2 is obtained".

- Page 6-51, line 4 – Should refer to the "Equation 6-19 can be modified…" or "Equation 6-21 can be modified…"

- Page 6-54, line 26 – Consider rewriting the sentence so to eliminate the use of both "following" and "followed".

- Page 6-55, line 16 – Should refer to "… gathering radon and progeny …"

- Page 6-55, line 32 – "More complicated systems…" is subjective. Is the measurement more complicated, the instrument more complicated, or is the data analysis more complicated? Is "sophisticated" better description than "complicated"? Where do charcoal canisters and gamma-ray measurements fit into the scheme, "simple" or "more complicated"?

- Page 6-56, Table 6.9 – The column "Time" is not clear; is that "total time", "sample collection time" or "count time"?

- Page 6-56, Table 6.9 – Regarding the "Remarks" column, "LLD" (lower limit of detection) should be replaced with "MDC". Also, the MDC listed are in a variety of different units which makes it difficult for the reader to compare/contrast the different techniques quickly and easily. Recommend an attempt be made to showing how MDC unit conversion should be performed while highlighting appropriate cautions and limitations. This table is another example where the use of citations to the scientific literature would greatly enhance the information contained in the table. With the citation, the reader has the option to quickly and easily getting more information on the method.

- Page 6-58, line 27 – Recommend replacing “surface barrier detector” with the much more common “ion-implanted planar silicon detector”.

- Page 6-59, line 25 – “sensitivity” should be replaced with “capability”.

- Page 6-63, Line 20 – Unclear what “averages 55,000 gamma in strength” is referring to. Best to specify in SI units (tesla, T) and include traditional units in paratheses.

Editorial comments on Appendix H:

- Page H-2, lines 12-13 – Recommend deleting “in the proper direction”.

- Page H-2, line 15 – Best to refer to “interchangeable detectors or probes” to use the common terminology.

- Page H-2, line 20 – “planchets” appears to be a typographic error.

- Page H-3, line 5 – Should correct “primary” and “secondary” to be consistent with the name of the probe.

- Page H-3, line 13 – Should replace “only” with “primarily”.

- Page H-3, line 29 – Should replace “only” with “primarily”.

- Page H-8, lines 16-17 – Should include a similar sentence for other instruments or delete from here.

- Page H-9, lines 7-8 – Should delete this sentence.

- Page H-9, line 13 – “same factor~~s~~”

- Page H-10, line 23 – Should refer to “… gamma-ray or x-ray radiation …”.

- Page H-11, line 3 – Should remove “alpha” from the secondary radiation.

- Page H-12, line 4 – Should refer to “Thallium-activated sodium iodide (NaI:Tl) …”.

- Page H-13, line 4 – Should refer to “Cerium-activated lanthanum bromide (LaBr:Ce) …”.

- Page H-15, line 20 – “24 hours” should be replaced with “72 hours”.

- Page H-15, line 30 – Recommend replacing “Detection capability” with “Detection sensitivity” so to be consistent with the rest of the appendix.

- Page H-24, line 8 – The phase “Dry cask storage neutrons” needs to be defined.

- Page H-51, Table H.1, column header, 3rd column – Should be “10^5” not “105” atoms.

- Page H-54, line 2 – Is this a typographical error? Should this be “Laboratory”? We are in the laboratory section.

- Page H-55, line 4 – Should be “lanthanides” not “lanthanum”.

- Page H-55, line 22 – Oxygen is also a major interferent that must be removed.

- Page H-55, line 27 – Missing a reference. Sentence reads “… has been reported by.”.

- Page H-59, Table H.2, last row, 4th column – Should read “requires P-10 gas”.

- Page H-60, Table H.2, 3rd and 4th rows – Consistently populate the table. For the gas-flow proportional counter “laboratory” and “field” are given separate rows, while liquid scintillation is combined and discussed on one row.

Editorial Comments from Charge Question 2.2:

- Page E-1, line 40 through page E-2, Line 2 – Recommend moving this paragraph to the beginning of Appendix E so that it is immediately available to the reader.

- Page E-6, line 6 – Need to insert a space between “median” and “of”.

- Page E-14, Example 3 – Two different notations ($\hat{\mu}_{\mathrm{RSS}}$, and $\overline{\mathrm{X}}_{\mathrm{RSS}}$) are used for the sample mean here. The document should use consistent notation to avoid confusion.

- In the footnotes for Tables E.1 – E.3, it looks as though the definitions for Δ and σ were inadvertently switched (it should be width of the gray region divided by the standard deviation, not the other way round).

- Section E.2.2 and Tables E.1 – E.3: Please give a clear statement of assumptions and also an attribution for these tables. If they were calculated in-house, please state the method. If they were taken from some other source, please cite it.

Editorial Comments from Charge Question 3.1:

- In Example 5 (p. 8-23), add a phrase such as "Based on the site survey measurements," before "the LBGR was selected to be…." to re-emphasize that this is site-specific and should not be expected to be an arbitrary or "cook-book" number.

- The same comment for p. 8-25, Example 6 and p. 8-30, Example 7.

- The first mention of the LBGR is on page 2-12, lines 12-19, under 'Overview of Radiation Survey and Site Investigation Process.' The section refers to Sections 5.3.3.1, 5.3.4.1, Appendix D, Section D.1.7.3 for more information; a reference to Chapter 4, Section 4.12.3.1 WRS Test should perhaps also be included here.

Editorial Comments from Response to Charge Question 3.3:

There are figures, text, and boxed text in the manual with incorrect references, references to sections that do not exist, or other logical errors, as follows:

- In both Figures 2-5 and 3-1, one of the steps asks, "Does Site Pose Immediate Risk to Human Health and Environment?" It appears the response choices to that question should be reversed; i.e., "Yes/Unknown" should lead to "Refer to Appropriate Regulatory Authority," and "No" should lead to the question about residual radioactivity. The same occurs in the next step, "Does Site Possibly Contain Residual Radioactive Material in Excess of Natural Background or Fallout Levels?" The responses to "Yes" and "No" should be reversed.

- In Figure 4.1, the box labeled "Identify Radionuclides" refers to Section 4.3, but Section 4.3 addresses survey types, not radionuclide identification.

- Section 5.3.5 refers to Section 4.2.5, but there is no Section 4.2.5.

- Appendix A.2.1 and A.2.2 do not refer to the appropriate sections in Chapter 4.

- In Appendix A, Figures A.1–A.9 are mentioned in the text, but the figure captions repeat A.1 and A.2 multiple times.

Editorial Comments from Response to Charge Question 3.4:

The SAB had the following editorial comments relating to Charge Question 3.4:

1. Section 5.1, Introduction should include a statement similar to the following: "Appendix O provides detailed calculations for statistical tests, illustrative examples for the determination of DCGLs, and more detailed derivations of key statical concepts and should be consulted when undergoing survey planning and design."

2. Section 5.3.3, The following statement or similar should be added: "Additional information on the WRS Test including an example is provided in Appendix O (O.2).
3. Section 5.3.4, The following statement or similar should be added: "Additional information on the Sign Test including an example is provided in Appendix O (O.3).
4. Section 5.3.5, The following statement or similar should be added: "Additional information on $DCLG_{EMC}$ including an example is provided in Appendix O (O.4).
5. Section 5.3.3.2, p. 5-31, when describing Table 5-2, text similar to the following should be added: "These values were calculated using Equation O-1 in Appendix O and increased by 20% to account for missing or unusable data and uncertainty in the calculated value of N."
6. Section 5.3.5 (p. 5-36, line 3): reference to Section 4.2.5 should be Section 4.5.2.
7. Section 5.3.5.2, Example 8, p. 5-41 states the following:

   "The distance between measurement locations for this number of data points and the given land area is 10 m, as illustrated in the application of Equation 5.1…. The grid area encompassed by a triangular sampling pattern of 10 m is approximately 86.6 m^2, as calculated…"

   In order to match the calculation of L in Example 8 that shows 10.7 m for a triangular grid, these statements should be changed have L= 10.7 m and the resulting grid area = 99.1 m^2, as follows:
   "The distance between measurement locations for this number of data points and the given land area is 10.7 m, as illustrated in the application of Equation 5.1…. The grid area encompassed by a triangular sampling pattern of 10.7 m is approximately 99.1 m^2, as calculated …"

APPENDIX C: ADDITIONAL COMMENTS

From Response to Charge Question 1.1

ANSI (1995) defines the acceptance criteria for verification testing by NIST as:

$$|V_R - V_N| < 3\sqrt{\sigma_R^2 + \sigma_N^2}$$

where

V_N	=	NIST value;
V_R	=	Reported value;
σ_N	=	1 sigma total uncertainty of V_N;
σ_R	=	1 sigma total uncertainty of V_R; and
$3\sqrt{\sigma_R^2 + \sigma_N^2}$	=	Measurement traceability limit for traceability claims.

The value of 3 in the formulation before the square root relates to a confidence level of approximately 99% in the comparison. A value of 2 would reduce the confidence level to 95%. The SAB advises EPA to select an appropriate value for MARSSIM scanning surveys.