Attachment E
Retrospective Power Assessment

THIS PAGE INTENTIONALLY LEFT BLANK

Retrospective Power Assessment

INTRODUCTION

Section 8.2.4 of the Multi-Agency Radiation Survey and Site Investigation Manual (MARSSIM) (Nuclear Regulatory Commission [NRC] 2000) states that an evaluation to determine whether data are consistent with the underlying assumptions made for statistical procedures helps to validate the use of a statistical test. One of the more important assumptions made in the final status survey (FSS) design for Site WR111 is that the number of samples collected in each survey unit (SU) are sufficient to achieve the data quality objectives (DQOs) set for the Type I (α) and Type II (β) error rates. Verification of the power of the statistical tests ($1-\beta$) to detect adequate remediation may be of particular interest. Methods for assessing retrospective power are in accordance with MARSSIM Appendix I.9.

POWER OF THE WILCOXON RANK SUM TEST

Section I.9.2 of Appendix I in MARSSIM provides guidance on calculating the power of the Wilcoxon Rank Sum (WRS) test. The minimum number of samples per SU calculated to meet the survey DQOs was 18. Therefore, SU boundaries were established such that a minimum of 18 systematic soil sample locations were included in each SU. The FSS design was set for Type I and Type II error rates of 0.05, providing a prospective power for the WRS test of 0.95, or 95 percent.

Following collection of the FSS data, updated estimates of statistical quantities describing the distribution of radioactivity within each SU became available. Retrospective power curves were generated for each SU based on the results of the FSS systematic soil sample analyses. The retrospective WRS power curves were calculated using the sum of ratios (SOR) data from systematic soil sample locations used to perform the WRS test. Table E.1 lists the statistical parameters used to construct the retrospective power curves. The retrospective power curves for SU1, SU2, SU3, and SU4 are shown in Figures E.1 through E.4, respectively.

Table E.1
WR111 Statistical Parameters Used to Construct Retrospective Power Curves

Survey Area	Mean SOR	Standard Deviation (SOR)	m (number of reference area samples)	n (number of survey unit samples)
Reference Area	2.18	0.16	18	NA
SU1	1.89	0.32	18	18
SU2	1.91	0.47	18	18
SU3	1.99	0.36	18	22
SU4	2.39	0.76	18	18

The WRS test was performed using SOR values; therefore, the derived concentration guideline level (DCGL) is 1.0 above background. The reference area mean was used as the most appropriate estimate of background, so a DCGL of 3.18 (the reference area mean SOR of 2.18 plus 1.0) was used for all four SUs, as shown on Figures E.1 through E.4.

The lower bound of the gray region (LBGR) for each SU was set equal to the mean SOR value for that SU. The mean SOR value was selected as the most appropriate estimate of the average level of radioactivity for a SU.

The standard deviation of the SOR values in each SU exceeded the standard deviation of the reference area SOR values, so the standard deviation of the SOR values was conservatively used to estimate the variability in each SU.

A cursory review of the mean SOR values in Table E.1 shows the mean SOR values for SU1, SU2, and SU3 are less than the mean SOR value in the reference area. This indicates there is almost no probability of deciding that the mean SOR value in one of these SUs exceeds the mean SOR in the reference area by more than the DCGL. The retrospective power curves in Figures E.1, E.2, and E.3 support this conclusion and show the retrospective power at the LBGR for SU1, SU2 and SU3, respectively, is effectively 1, or 100 percent.

SU4 has a mean SOR value of 2.39, which is greater than the reference area mean SOR of 2.18. The standard deviation for the SOR values in SU4 is 0.76. The DCGL (3.18) minus the LBGR for SU4 (2.39) gives a shift of 0.79. The shift was divided by the standard deviation in SU4 (0.76), resulting in a relative shift of 1.04. The SU4 relative shift value is smaller than the relative shift value of 1.6 used to design the final status survey. The retrospective power curve in Figure E.4 shows the power at the mean SOR value in SU4 is less than 0.95, indicating the potential for making a Type II decision error is higher than what was planned.

The power calculation for the WRS test in MARSSIM Appendix I depends on values of two parameters called P_r and p_2 in MARSSIM Appendix I. Table I.10 in MARSSIM lists values of P_r and p_2 for select values of the relative shift. Table I.10 provides values for a relative shift of 1.0 and a relative shift of 1.1. Linear interpolation was used to calculate the value of P_r for the actual relative shift of 1.04 for SU4.

$$\frac{1.1 - 1.0}{1.01 - 1.0} = \frac{P_r(1.1) - P_r(1.0)}{P_r(1.04) - P_r(1.0)}$$

$$2.5 = \frac{P_r(1.1) - P_r(1.0)}{P_r(1.04) - P_r(1.0)}$$

$$P_r(1.04) - P_r(1.0) = \frac{P_r(1.1) - P_r(1.0)}{2.5}$$

$$P_r(1.04) = P_r(1.0) + 0.4(P_r(1.1) - P_r(1.0))$$

$$P_r(1.04) = 0.760250 + 0.4(0.781662 - 0.760250)$$

$$P_r(1.04) = 0.768815$$

Linear interpolation was also used to calculate the value of p_2 for the actual relative shift of 1.04 for SU4.

$$p_2(1.04) = p_2(1.0) + 0.4(p_2(1.1) - p_2(1.0))$$

$$p_2(1.04) = 0.633702 + 0.4(0.662216 - 0.633702)$$

$$p_2(1.04) = 0.645108$$

The mean of the Mann-Whitney form of the WRS test statistic ($E(W_{MW})$) is calculated using Equation I-9 from MARSSIM Appendix I.

$$E(W_{MW}) = mnP_r$$

$$E(W_{MW}) = 18 \times 18 \times 0.768815 = 249.096$$

Where:

m	=	number of reference area measurements (18 for SU4)
n	=	number of survey unit measurements (18 for SU4)

The variance of the Mann-Whitney form of the WRS test statistic ($Var(W_{MW})$) is calculated using Equation I-10 from MARSSIM Appendix I.

$$Var(W_{MW}) = mnP_r(1 - P_r) + mn(n + m - 2)(p_2 - P_r^2)$$

$$Var(W_{MW}) = (18 \times 18 \times 0.768815 \times 0.231185) +$$
$$(18 \times 18 \times 34 \times (0.645108 - 0.591077)$$

$$Var(W_{MW}) = 57.5873 + 595.205 = 652.79$$

The power of the WRS test is computed using Equation I-8 from MARSSIM Appendix I.

$$Power = 1 - \Phi\left(\frac{W_c - 0.5 - 0.5m(m+1) - E(M_{MW})}{\sqrt{Var(M_{MW})}}\right)$$

$$Power = 1 - \Phi\left(\frac{385 - 0.5 - 171 - 249.096}{25.55}\right) = 1 - \Phi(-1.393) = \Phi(1.393)$$

Where:

W_c	=	critical value of the WRS test (MARSSIM Table 1.4, 385 for m=18, n=18, α=0.05 for SU4)
Φ	=	standard normal cumulative distribution function

MARSSIM Table I.1 provides values for the standard normal cumulative distribution function. The value of the function for 1.39 is 0.9177, and the value for 1.40 is 0.9192. The value for SU4 falls between these two values in Table I.1. Therefore, the power of the WRS test for SU4 is 0.92, or 92 percent.

The probability of deciding to continue to remediate SU4 when SU4 demonstrated compliance with the 25 millirem per year release criterion was approximately 8 percent instead of the 5 percent that was planned. Portions of SU4 had already been remediated twice. Thorium-230 contributes significantly to the average SOR and cannot be identified using scanning techniques, requiring additional resources and making real-time remedial action decisions more difficult. The number of systematic samples was adequate based on the initial survey design, but the presence of thorium-230 made complete remediation difficult to determine, resulting in the variability in SU4 sample concentrations that was higher than originally planned.

In sum, we used the guidance in MARSSIM Appendix I to generate the retrospective power curves for each SU at Site WR111. Retrospective power curves shown in Figures E.1, E.2, and E.3 show the retrospective power at the LBGR for SU1, SU2 and SU3, respectively, is effectively 1, or 100 percent. Using Equations I-8, I-9, and I-10 from MARSSIM Appendix I, the retrospective power for SU4 at the mean SOR was calculated to be 0.92. The number of samples

in SU4 is sufficient to support decisions regarding compliance with a Type II decision error rate of 0.1 instead of 0.05 which was originally planned. The consequences of a Type II decision error would result in additional remediation that may not produce a reduction in total effective dose equivalent. Therefore, it is concluded that the number of samples in each SU is sufficient to meet the project DQOs, and no further action is recommended at WR111.

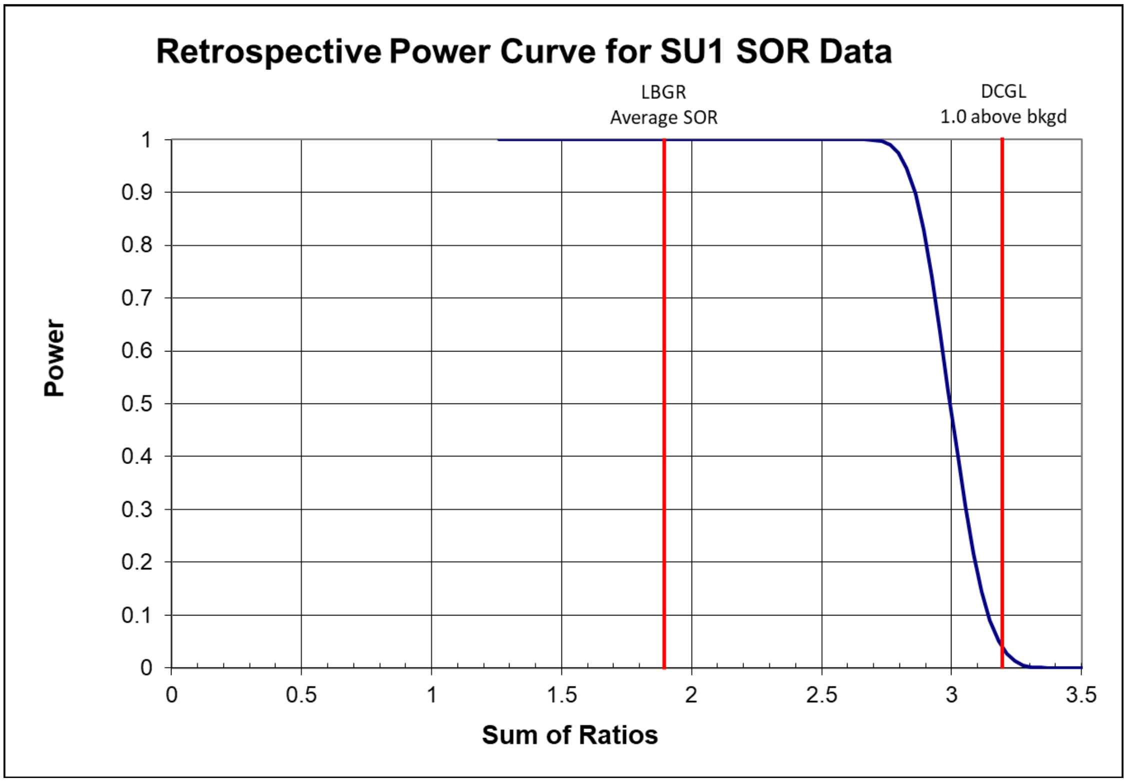

Figure E.1 SU1 Retrospective Power Curve

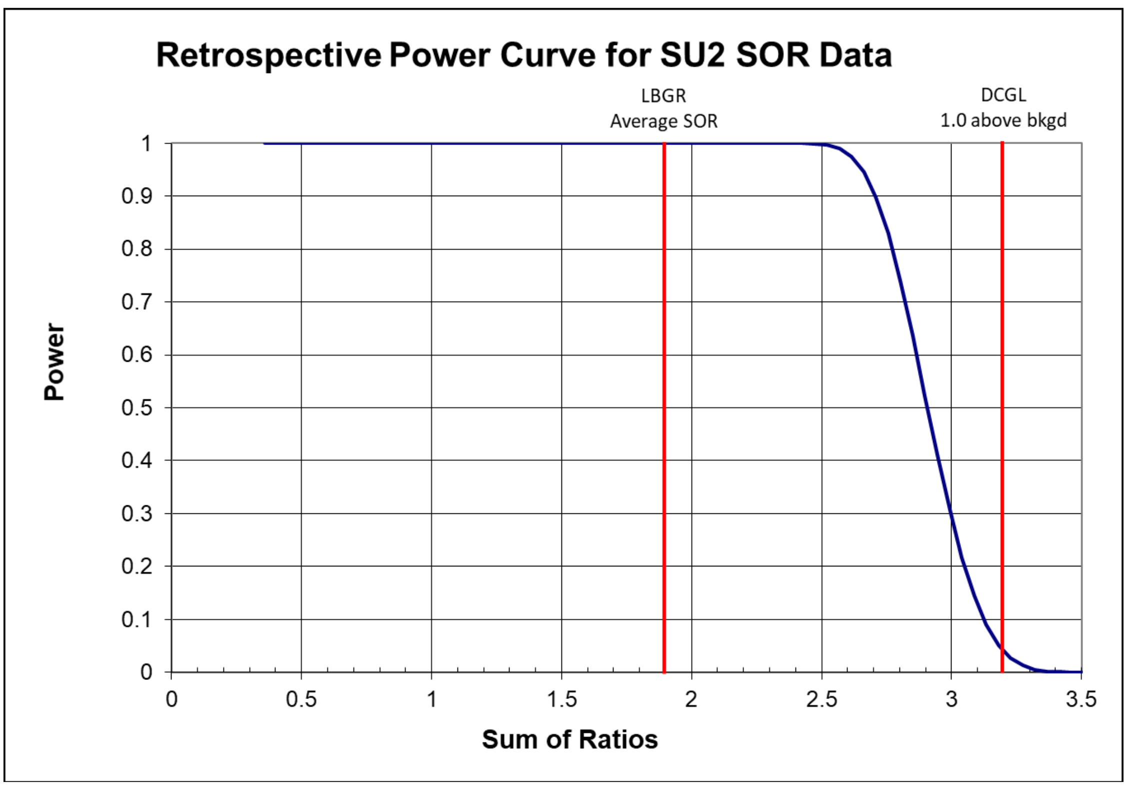

Figure E.2 SU2 Retrospective Power Curve

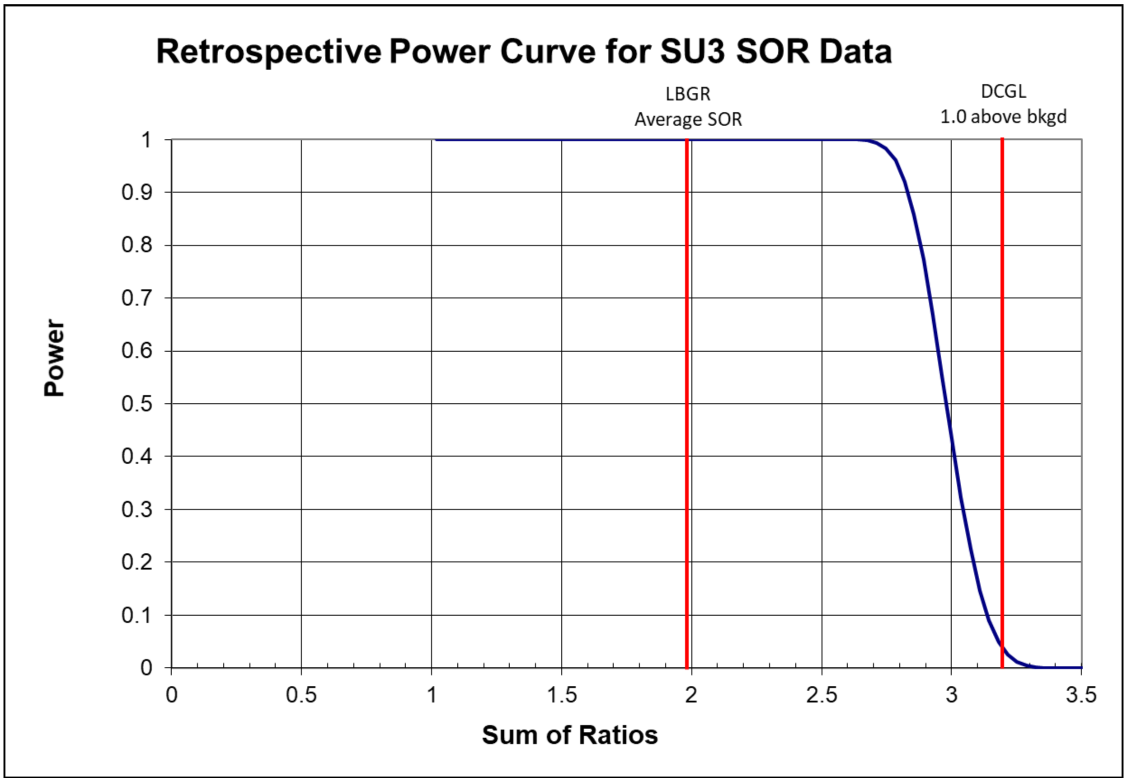

Figure E.3 SU3 Retrospective Power Curve

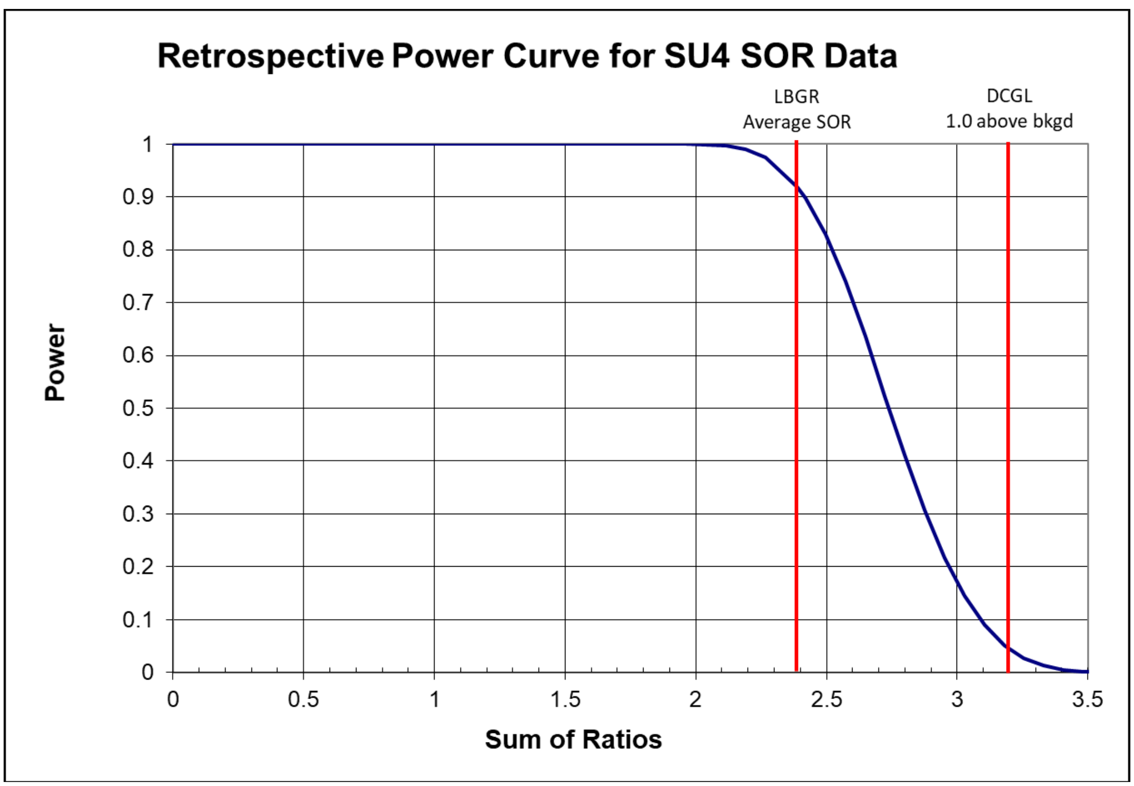

Figure E.4 SU4 Retrospective Power Curve