Increasing the Accuracy of the 99th Percentile of PC

White Paper

National Institute for Occupational Safety and Health (NIOSH)

July 2022

Daniel O. Stancescu
Division of Compensation Analysis and Support, NIOSH

A. Iulian Apostoaei, Brian A. Thomas
Oak Ridge Center for Risk Analysis, Inc. (ORRISK, Inc.)

Reviewed by Timothy D. Taulbee
Division of Compensation Analysis and Support, NIOSH

ABSTRACT

Claims filed under the Energy Employees Occupational Illness Compensation Program Act (EEOICPA) are awarded compensation when the estimated **p**robability of cancer **c**ausation (PC) for exposed workers exceeds the decision threshold of 50%. NIOSH-IREP estimates PC as a probability distribution function (pdf) from multiple samples produced using Monte-Carlo iterations to account for uncertainties in radiation dose and risk parameters representing the exposure situation, and the 99th percentile of that distribution is used in decision making. Currently, 2,000 samples are used in IREP to determine the 99th PC values. When these values are between 45% and 52%, 30 runs of 10,000 iterations each are carried out using IREP Enterprise Edition (IREP-EE) and the average of the resulting thirty 99th percentiles (Avg. 99th PC value) is used for compensation.

This report studies ways to further increase precision of the estimated 99th percentile of PC, when expected values are close to the threshold of 50%. As such, this report investigates:
(a) the convergence of the 99th percentile of PC as the number of iterations is increased in IREP,
(b) different settings that can used to achieve a desired precision for the 99th percentile of PC,
(c) the use of the IREP Predictive Tool to determine the minimum number of iterations necessary to achieve a given precision. Analyses include validation tests between the regular and EE versions of IREP.

A total 23 cases were selected for testing. These cases had an Avg. 99th PC values between 49.00% and 50.01% based on the current NIOSH-IREP method involving 30 runs at 10,000 iterations. The convergence of the 99th PC values as well as of the Avg. 99th PC values, as the number of iterations increases was shown for multiple cases with PC values between 49.00% and 50.10%, and for each of these cases there is a clear trend for both quantities to converge to a stable value that represents the theoretical 99th percentile from the distribution of PC values for that case.

The theoretical 99th PC value of the PC probability distribution for a case cannot be determined exactly, but a 95% confidence interval for the estimated 99th PC value can be estimated. The larger the number of iterations used in the computation, the smaller the length of the 95% confidence interval for the 99th PC value will be, where the length is defined as the upper bound minus the lower bound of the 95% confidence interval. For a sample size large enough, a length of the 95% confidence interval of less than 0.15 or 0.10 can be achieved.

The 95% confidence intervals for the Avg. 99th PC values with a length lower than 0.15 were obtained in IREP-EE for all tested cases when using 200 runs at 50,000 iterations. Lengths of the 95% confidence intervals for the Avg. 99th PC values lower than 0.1 were obtained for all tested cases when using 300 runs at 100,000 iterations.

To minimize running time, an IREP Predictive Tool was developed to determine the minimum the number of iterations needed to achieve a desired precision measured by the length of the 95% confidence interval of the Avg. 99th PC values. For a 0.15 precision level, the number of

iterations needed for the 23 cases with Avg. 99th PC values between 49.00% and 50.01% varied between 5,000 and 35,000 iterations, and for the 0.10 precision level, the number of iterations needed for the 23 cases varied between 15,000 and 60,000 iterations. The largest IREP running time observed for the 23 cases was less than one hour for a precision level of 0.15 and less than 1.75 hours for a precision level of 0.10.

By increasing the number of runs in IREP-EE from 30 runs to 300 runs, and by increasing the number of iterations per run to values predicted by IREP Predictive Tool, the length of the 95% confidence interval of the Avg. 99th PC values can be reliably reduced below one of the desirable precision levels of 0.15 or 0.10.

TABLE OF CONTENTS

1.0 INTRODUCTION

This document follows the "Effect of Alternative Percentile Definition on PC values" white paper (Stancescu et al., 2021) and investigates: (a) the convergence of the 99th percentile of PC as the number of iterations is increased in IREP, (b) the different settings that can used to achieve the desired precision for the 99th percentile of PC, (c) the use of the IREP Predictive Tool to minimize the number of iterations, and (d) the validation of the 99th percentiles of PC from the regular and EE versions of IREP.

2.0 PC VARIABILITY ACROSS RANDOM SEEDS

The NIOSH-IREP software uses Monte-Carlo simulations to estimate the distribution of PC values for a case, and the 99th percentile of this distribution is used to adjudicate the case, by comparing the 99th percentile with the 50.0% compensation threshold.

In order to use the Monte-Carlo simulation technique, a random seed value must be specified to generate a unique sequence of pseudorandom numbers that are used in sampling multiple probability distributions, which are combined during the computation process to obtain the final distribution of the PC values. For this reason, the chosen random seed uniquely determines the final 99th PC value. If a different random seed is chosen, the final 99th PC value will vary slightly, depending also on the number of iterations used (i.e., sample size), and other factors specific to each case (cancer model, uncertainty associated with radiation doses, etc.). The fact that the random seed value determines the final 99th PC value has the advantage of the 99th PC value being reproducible, which is very important in the last step of the adjudication process that needs to be performed by the Department of Labor. However, this also raises the question of which result is valid, if different random seeds are used and the answer to that question is that all the results are equally valid estimates. To avoid the issue of picking a more favorable random seed (a random seed that would produce a higher 99th PC value), or a less favorable random seed (a random seed that would produce a lower 99th PC value), the random seed is set the same for all cases, i.e. it is chosen to be the number 99 for all cases run on the IREP version.

A slightly different procedure is applied for cases run on the IREP Enterprise Edition (IREP EE), which is used for cases that have the 99th PC value (obtained from the IREP version) ranging between 45.00% and 51.99%. The current procedure is that the case is rerun on IREP EE 30 times with 10,000 iterations, where each of the 30 runs uses a different random seed value. This procedure averages the 99th PC values obtained from the 30 runs, which has the advantage of eliminating the influence of the extreme 99th PC values (values either too low or too high) that might be generated during this process. The resulting average 99th PC value is much more stable to the effect of changing the whole set of 30 random seeds, compared with individual 99th PC values, which have a much greater variability when the random seeds are changed.

The validation between the 99th PC values obtained from the IREP version using one run with a large number of iterations versus the Avg. 99th PC values obtained from the IREP Enterprise Edition using multiple runs with smaller number of iterations is addressed in a later section.

3.0 PC CONVERGENCE AT 30 RUNS IN IREP ENTERPRISE EDITION

The convergence of the 99th PC values in IREP and the convergence of the Avg. 99th PC values in IREP EE, as the number of iterations increases, can be seen in Figure 1, which shows the 99th PC values obtained for a case with the 99th PC value on record between 49% and 50% (the Avg. 99th PC value on record in NOCTS for this case is 49.87%). The plot shows the individual 99th PC values, as well as the Avg. 99th PC values from 30 runs, obtained at 7 different sample sizes ranging between 10000 and 200000 iterations. For each sample size, four sets of 30 random seeds were used (the same 4 sets of 30 random seeds were used at each sample size). The plot clearly shows that for each sample size, there is much more variation in the individual 99th PC values than in the Avg. 99th PC values obtained from each set of 30 random seeds. Also, as the sample size increases, the Avg. 99th PC values obtained are much more stable with respect to changing the set of 30 random seeds. There is a clear trend of the individual 99th PC values, as well as the Avg. 99th PC values, to converge to a stable value that represents the theoretical 99th PC value of the distribution of PC values for this case.

The theoretical 99th PC value of the distribution of PC values for a case cannot be determined exactly, but a 95% confidence interval can be estimated for the theoretical 99th PC value. The larger the number of iterations used in the computation, the smaller the length of the 95% confidence interval for the theoretical 99th PC value will be. This can be seen in Figure 2, which shows the 95% confidence intervals for the theoretical 99th PC value obtained based on a 1,000,000-iteration sample (purple dotted lines), as well as based on a 5,000,000-iteration sample (blue dotted lines), for the same case shown in Figure 1. The 95% confidence interval for the theoretical 99th PC value based on 1,000,000 iterations is (49.63, 50.02), and the 95% confidence interval for the theoretical 99th PC value based on 5,000,000 iterations is (49.76, 49.93). As the sample size increases, the 95% confidence interval for the theoretical 99th PC value will be narrower. In this case, the length of the 95% confidence interval for the theoretical 99th PC value obtained from the 1,000,000-iteration sample is 0.39, and the length of the 95% confidence interval for the theoretical 99th PC value obtained from the 5,000,000-iteration sample is 0.17, and these lengths remain almost unchanged if different random seeds are used. The green dotted line shows the Avg. 99th PC value on record in NOCTS for this case, which is 49.87%, and lies inside both 95% confidence intervals for the theoretical 99th PC value.

Figure 3 shows the 95% confidence intervals for the Avg. 99th PC values at each sample size, obtained using both the default and the alternative percentile definitions (Stancescu et al. 2021), as well as the 95% confidence limits for the theoretical 99th PC value obtained based on a 5,000,000 sample (red dotted lines), for the same case shown in Figures 1 and 2. The Avg. 99th PC value obtained based on the alternative percentile definition is always greater or equal than the Avg. 99th PC value obtained based on the default percentile, but the differences between the Avg. 99th PC values obtained using the two percentile definitions get smaller as the sample size increases. The 95% confidence intervals around the Avg. 99th PC values decreases as the sample size is increased. The lengths of the 95% confidence intervals for the Avg. 99th PC values at 200,000 iterations are smaller or equal than 0.17, which is the length of the 95% confidence interval for the theoretical 99th PC value obtained from a 5,000,000 sample.

Figure 4 shows the 95% confidence intervals for the Avg. 99th PC values at each sample size, obtained using both the default and the alternative percentile definitions, as well as the 95% confidence limits for the theoretical 99th PC value (red dotted lines), for a different case with the 99th PC value on record slightly above 50% (the Avg. 99th PC value on record in NOCTS for this case is 50.01%). The 95% confidence interval for the theoretical 99th PC value for this case was obtained by using a non-parametric method (Meeker 2017) from a data set with 6,000,000 individual PC values. This data set was obtained by combining 3 groups of 2,000,000 individual PC values, since IREP cannot directly compute the 95% confidence limits for a case with large number of exposures and for a large sample size, due to computer memory limitations. The 95% confidence interval for the theoretical 99th PC value is (49.96, 50.05), which shows that for this particular case, even with a large sample size, the compensability of the case might be hard to decide. A similar situation is shown in Figure 5, for a case with the 99th PC value on record slightly lower than 50% (the Avg. 99th PC value on record in NOCTS for this case is 49.90%). The 95% confidence interval for the theoretical 99th PC value for this case was obtained by using the same non-parametric method for a data set with 10,000,000 individual PC values, which was obtained by combining 10 samples with 1,000,000 individual PC values. The 95% confidence interval for the theoretical 99th PC value is (49.95, 50.07), which shows that for this case too, even with a very large sample size, the compensability of the case might be hard to decide. The 95% confidence intervals for the theoretical 99th PC values for these two cases are almost centered around the 50.0% compensation threshold, which suggests that the compensability of such unusual cases might be hard to decide.

Plots showing the convergence of the Avg. 99th PC values from 30 runs, as the sample size increases, for 11 additional cases with Avg. 99th PC values between 49.00% and 50.10%, are shown in Appendix A. The 95% confidence interval for the theoretical 99th PC value for these cases were obtained at 5,000,000 iterations.

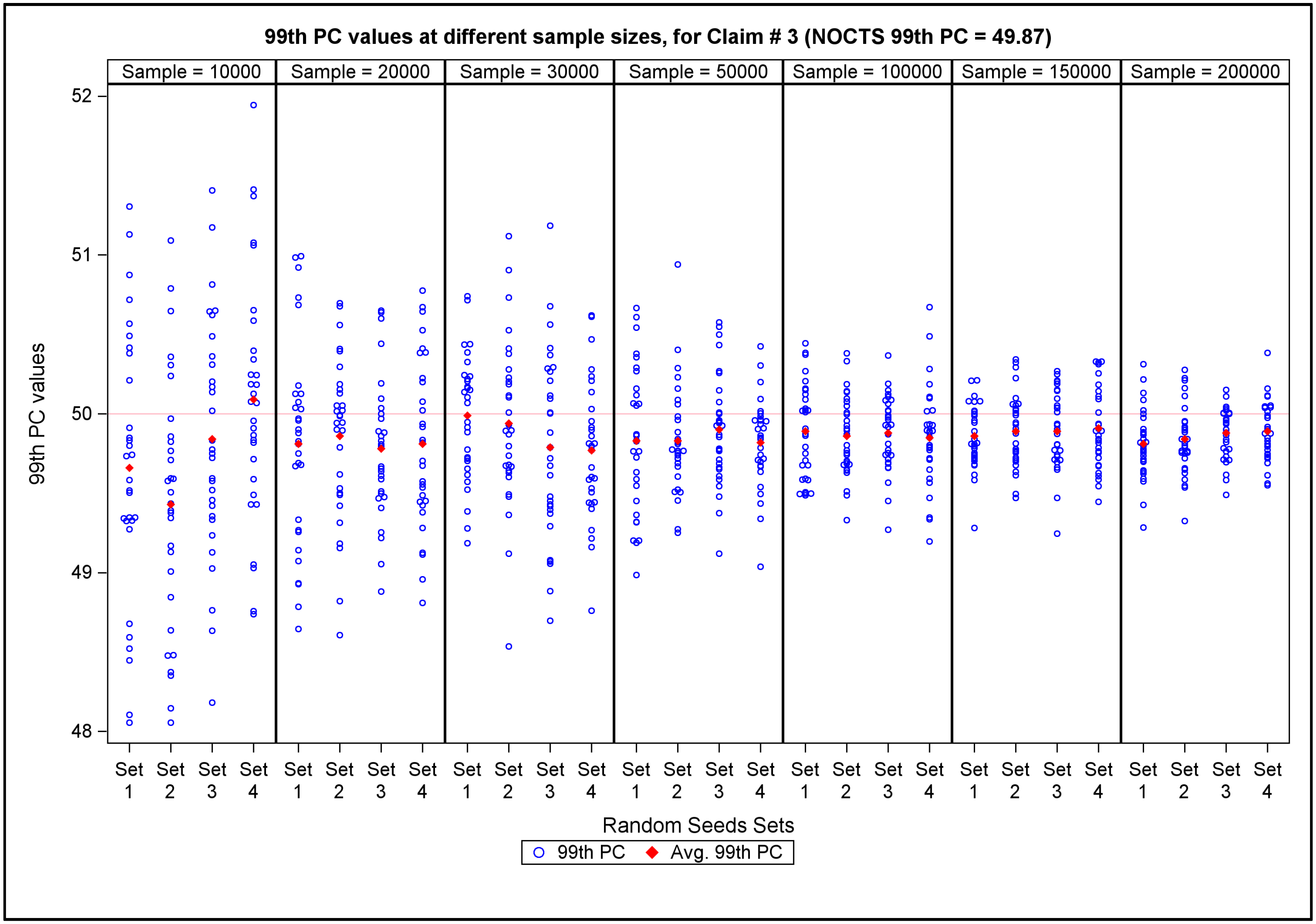


Figure 1. The 99th PC values obtained at different sample sizes, for a case with the Avg. 99th PC value between 49% and 50%.

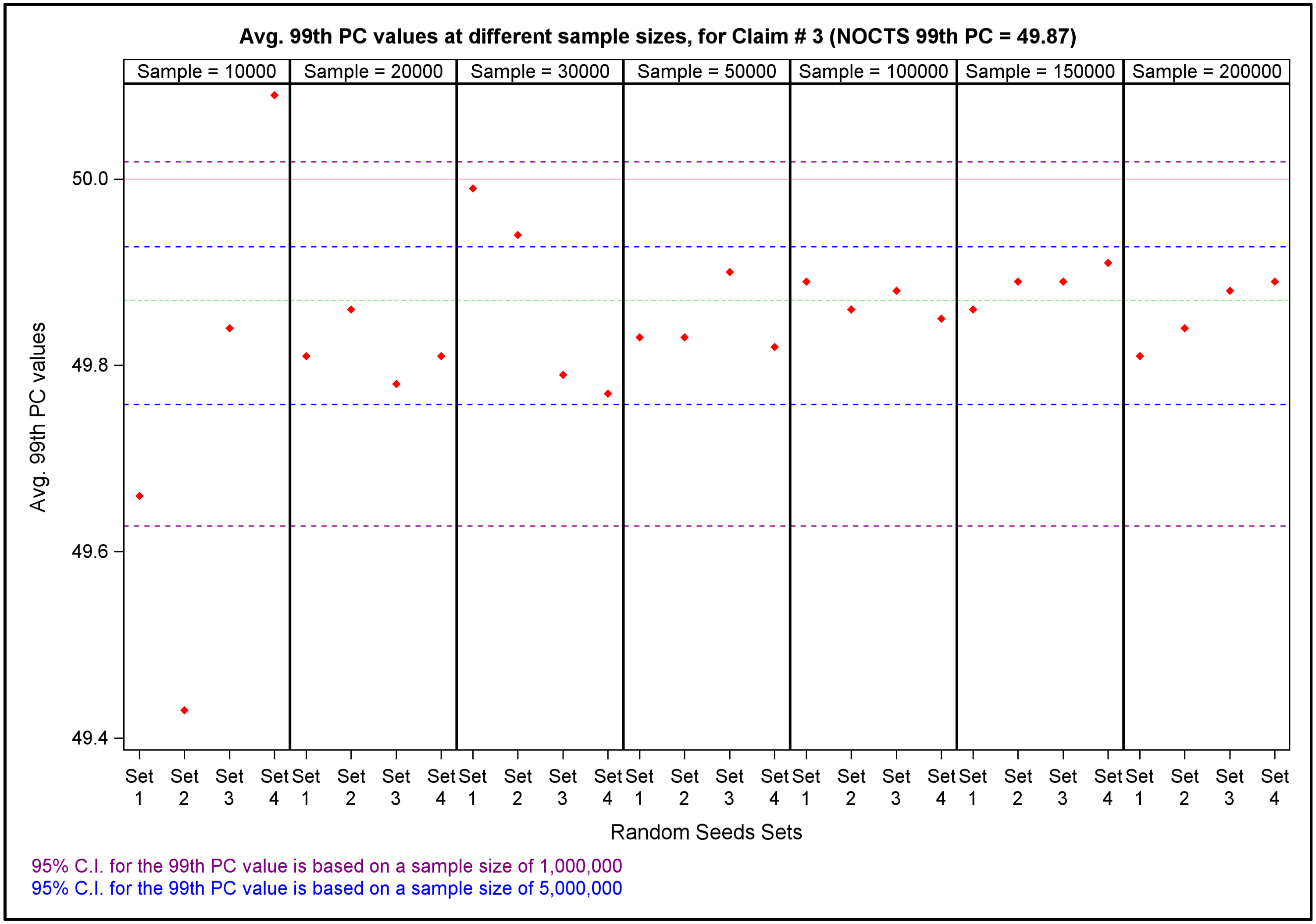


Figure 2. The Avg. 99th PC values from 30 runs, and the 95% confidence limits for the theoretical 99th PC value, obtained using two different sample sizes.

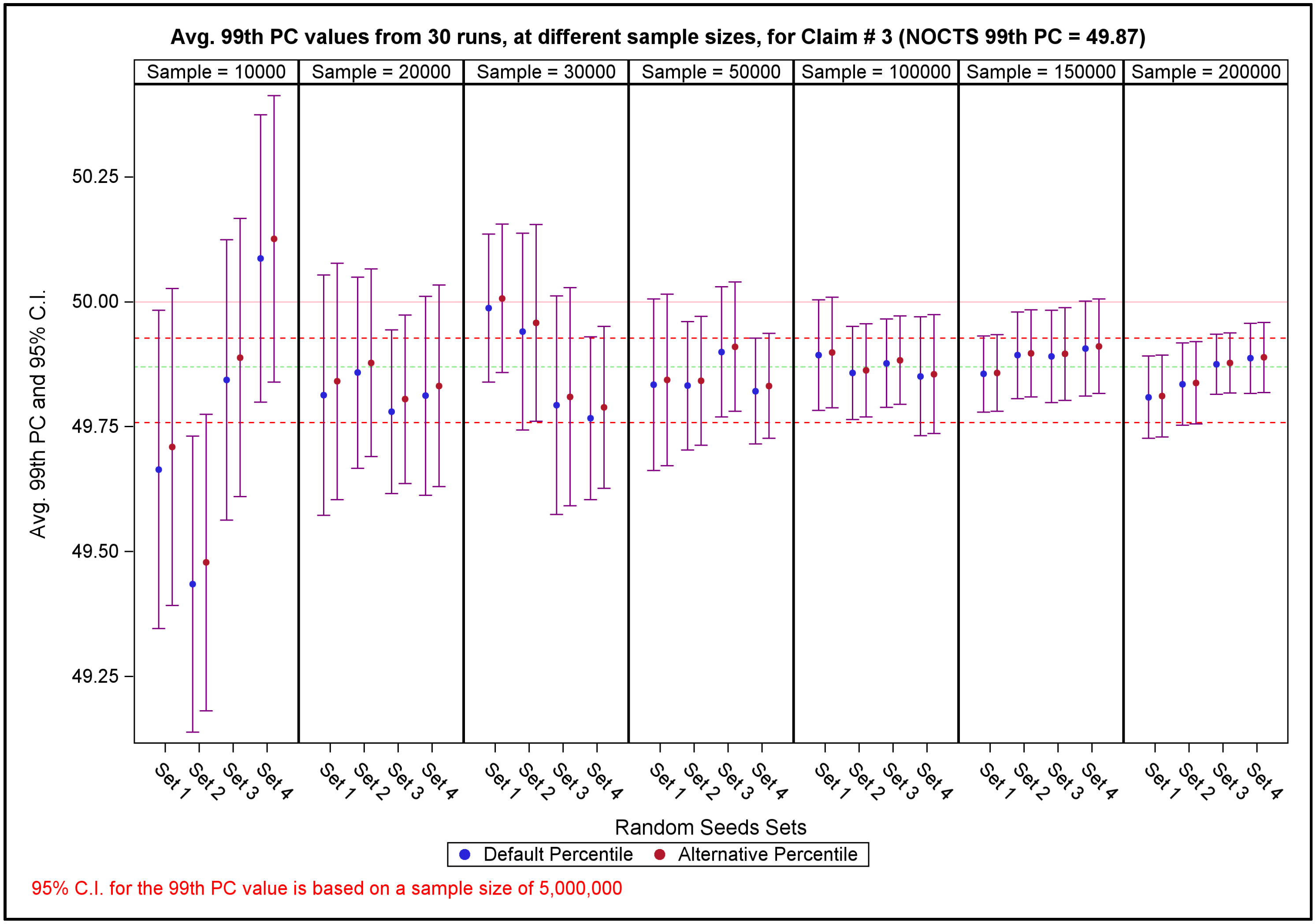


Figure 3. The Avg. 99th PC values from 30 runs, and their 95% confidence intervals, obtained using the default and the alternative percentiles.

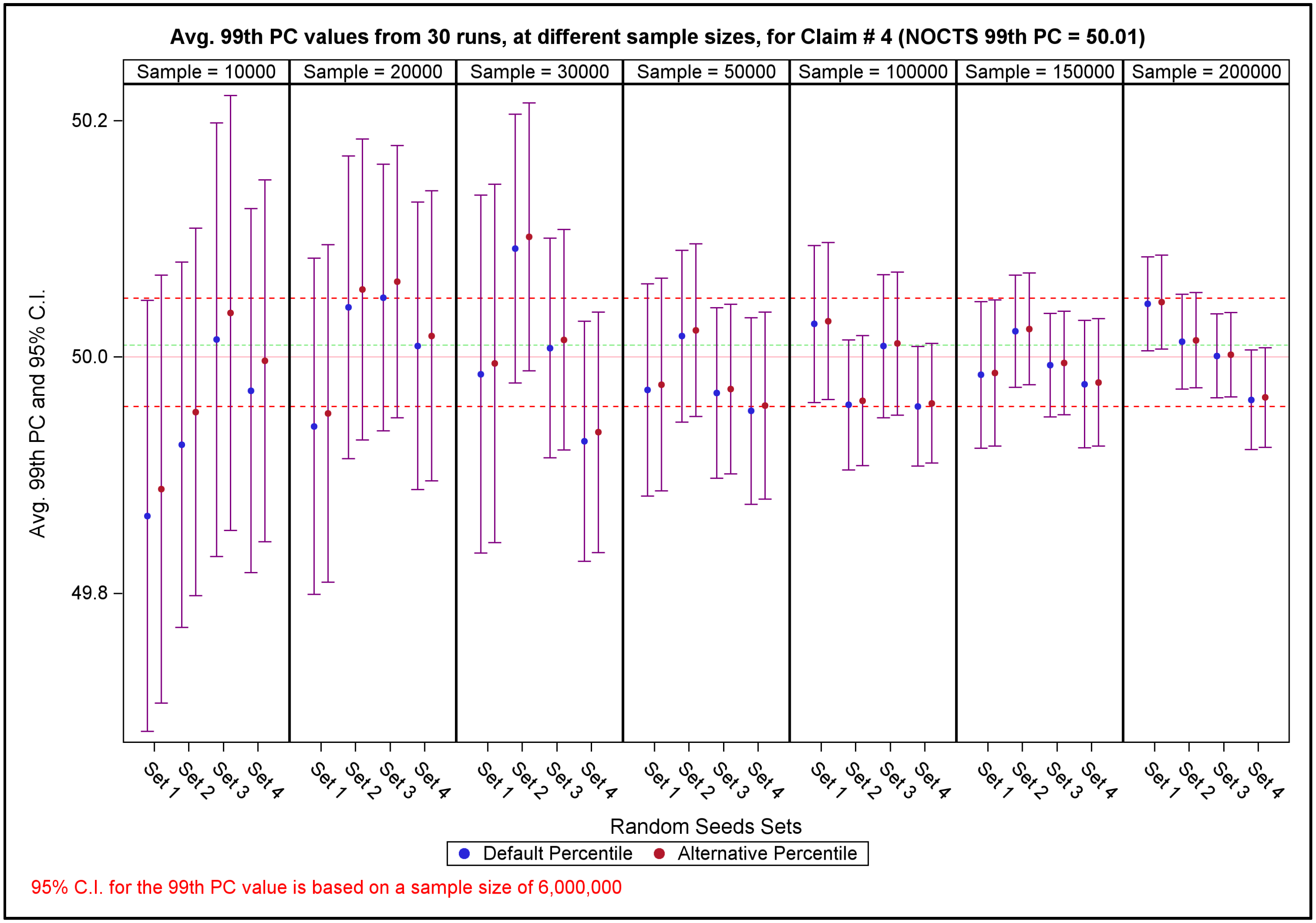


Figure 4. The Avg. 99th PC values from 30 runs, and their 95% confidence intervals, obtained using the default and the alternative percentiles.

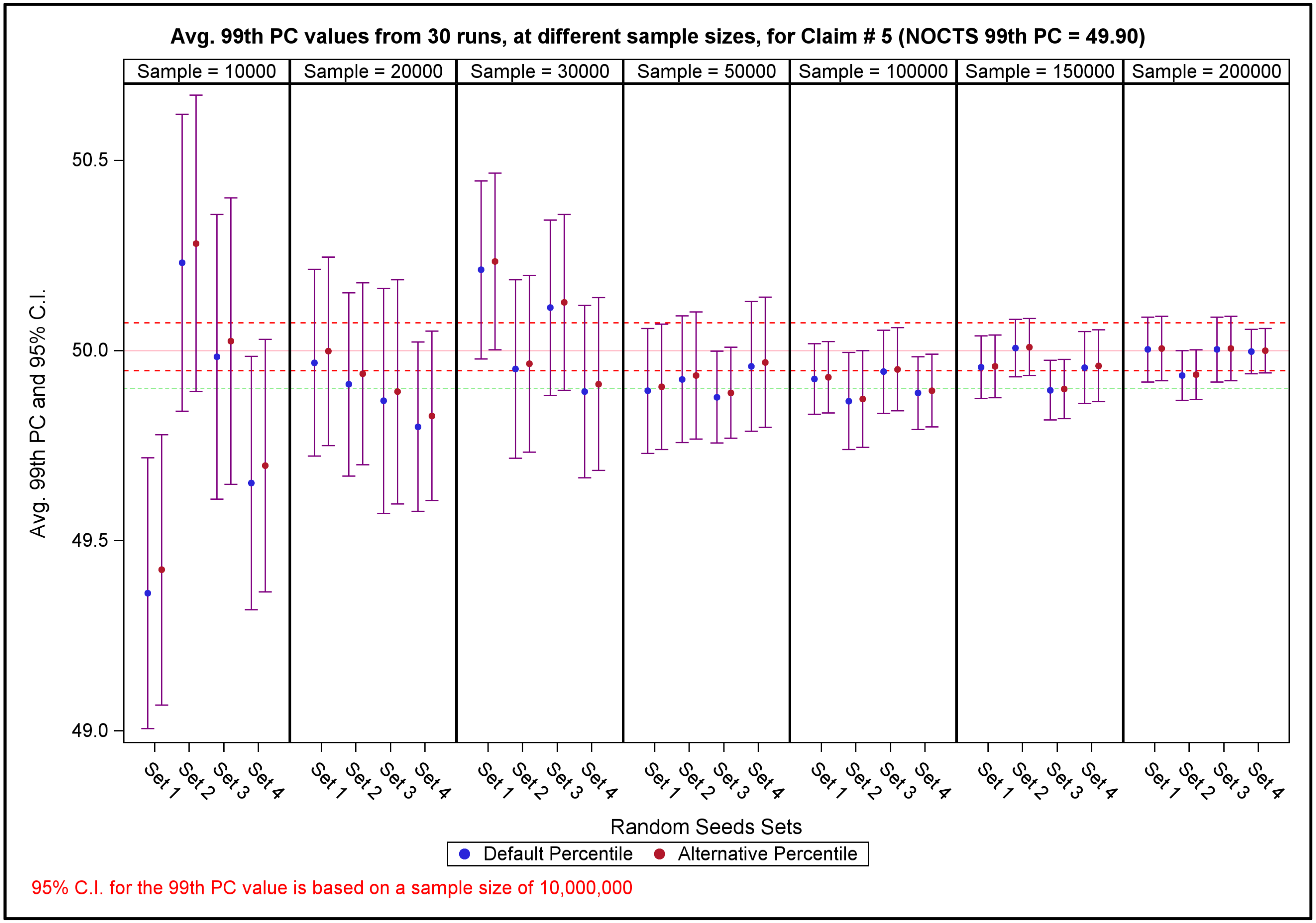


Figure 5. The Avg. 99th PC values from 30 runs, and their 95% confidence intervals, obtained using the default and the alternative percentiles.

4.0 PC CONVERGENCE AT MORE THAN 100 RUNS IN IREP EE

In the previous section, it was shown that the Avg. 99th PC values converge to a stable value that represents the theoretical 99th PC value of the distribution of PC values for that case. In order to achieve convergence for the Avg. 99th PC values (i.e., the Avg. 99th PC values to become relatively stable with respect to changing the set of random seeds), the number of iterations needs to be increased to at least 50,000 iterations for each of the 30 runs used in the current version of IREP Enterprise Edition, for most of the tested cases. To speed up the convergence of the PC values and to reduce the length of the 95% confidence intervals for the Avg. 99th PC values, either the number of iterations needs to be increased or the number of runs needs to be increased.

The goal is to reduce the 95% confidence intervals for the Avg. 99th PC values to be less than or equal to a desired precision level (e.g., a length of the 95% confidence interval of 0.10 or 0.15), for cases with Avg. 99th PC values in the 49.00% to 49.99% range. Increasing the number of iterations is not always a good option, especially for cases with larger number of exposures, since the IREP runs can get interrupted due to the high memory usage. The option of increasing the number of runs is more feasible in practice, since it just adds additional runs to the 30 runs previously used in the IREP Enterprise Edition. To look for the optimal combination of the number of runs and iterations that are enough to reduce the length of the 95% confidence intervals for the Avg. 99th PC values below the desirable lengths of either 0.10 or 0.15, several different combinations of number of runs and iterations were considered. Another factor in choosing the optimal combination of the number of runs and iterations is the total time it takes in IREP EE to compute the Avg. 99th PC value for each combination of runs and iterations; ideally, a single cancer case should be completed in no more than 2 or 3 hours.

In this analysis we selected from NOCTS the 22 single cancer cases with the Avg. 99th PC values between 49.00% and 49.99%, and one single cancer case with the 99th PC value just slightly above the 50.0% compensation threshold (Avg. 99th PC value on record for that case is 50.01%). These 23 cases with the Avg. 99th PC values between 49.00% and 50.01% were run on a modified version of the IREP Enterprise Edition for the following four scenarios:

- Scenario 1: 100 runs with 50,000 iterations (for a total of 5,000,000 iterations)
- Scenario 2: 200 runs with 50,000 iterations (for a total of 10,000,000 iterations)
- Scenario 3: 300 runs with 50,000 iterations (for a total of 15,000,000 iterations)
- Scenario 4: 300 runs with 100,000 iterations (for a total of 30,000,000 iterations).

Figure 6 shows a plot with the IREP running times obtained from the four scenarios versus the number of exposures for all 23 cases with Avg. 99th PC values between 49.00% and 50.01%. The largest number of exposures among the 23 cases is 380. The filled circles in the plot correspond with cases of lung cancers, which usually take longer to complete due to the two lung cancer models that are run in parallel. The IREP running times were obtained on a new research server, which has a very fast processor and 128 GB of RAM memory.

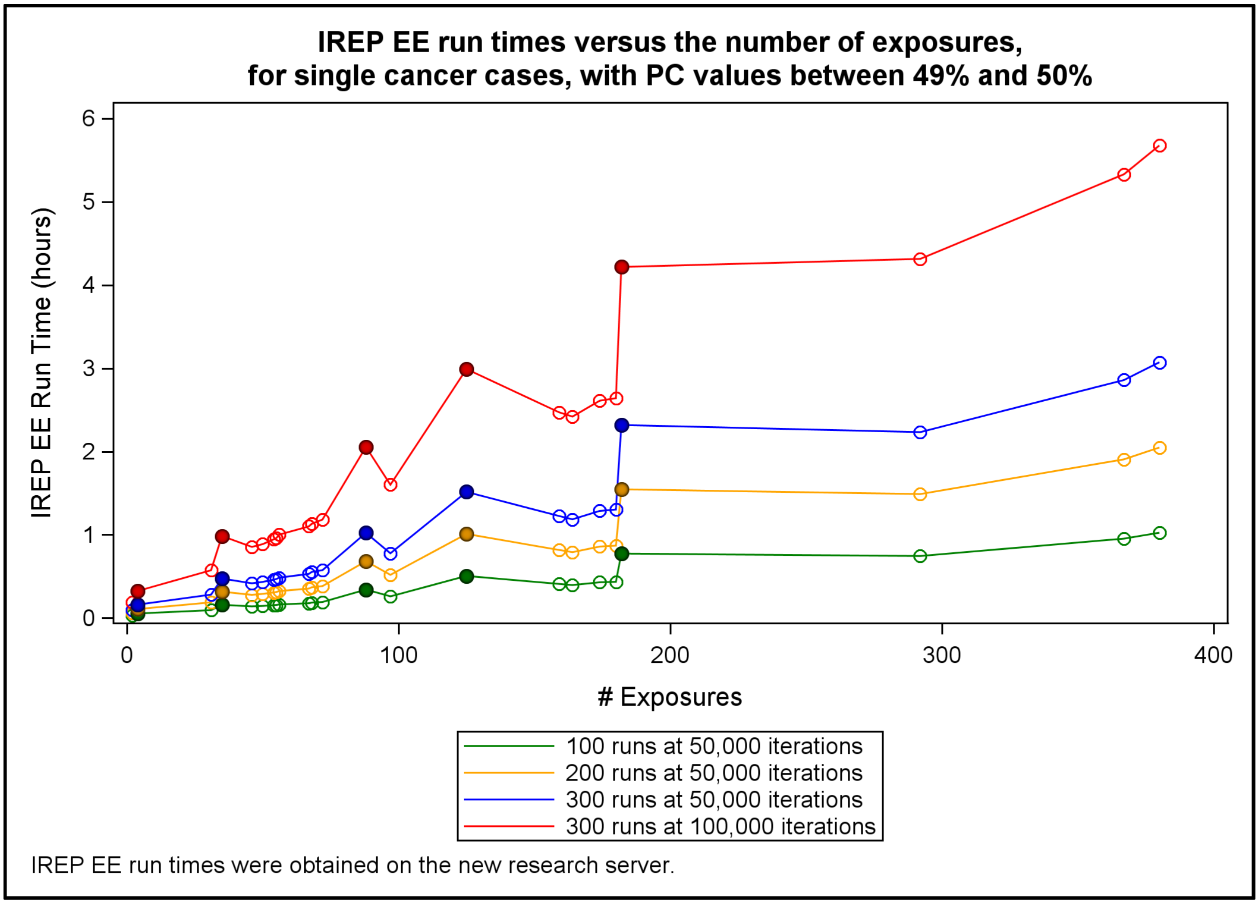


Figure 6. IREP EE running times for the 23 single cancer cases with Avg. 99th PC values between 49.00% and 50.01%.

The largest IREP running time observed for the 23 cases using Scenario 1 is around 1 hour, using Scenario 2 is around 2 hours, using Scenario 3 is around 3 hours, and using Scenario 4 is slightly less than 5.7 hours.

The lengths of the 95% confidence intervals obtained for each of the 23 cases from the four scenarios with different number of runs and iterations are shown in Figures 7 through 10. Reference lines for the 0.10 and 0.15 desired precision lengths of the 95% confidence intervals are shown in the plots. For each case, the length of the 95% confidence intervals is highest in Scenario 1, and then decreases in Scenarios 2, 3, and 4, with Scenario 4 having the smallest length of the 95% confidence interval for each case. The length of the 95% confidence intervals depends on the characteristics of each case (cancer model, number of exposures, uncertainty in the doses), but the trend of these 95% confidence intervals lengths is to get smaller as the total number of iterations used is increased.

Figure 11 summarizes the lengths of all the 95% confidence intervals obtained for the 23 cases from each of the four scenarios, and it also shows the standard errors for each case obtained from the four scenarios (the length of the 95% confidence intervals as well as the standard errors were obtained from a bootstrap analysis). The plot shows that the lengths of the 95% confidence intervals are smaller than 0.15 for all cases in Scenarios 2 and 3, and less than 0.10 for all cases in Scenario 4. For Scenario 3, the lowest length of the 95% confidence intervals is 0.05, and for Scenario 4, the lowest length of the 95% confidence intervals is 0.03; however, these very small lengths for the 95% confidence intervals can only occur in rare conditions when the cancer model or the total dose assigned to a case have distributions with very small uncertainties.

The lengths of the 95% confidence intervals obtained for 13 cases from 11 scenarios with different numbers of runs and iterations (the 4 scenarios with 100, 200, or 300 runs, and 7 scenarios with 30 runs) are shown in Figures 12 through 16. These plots show that by using only 30 runs (as is currently done in the IREP Enterprise Edition), even with 200,000 iterations per run, there will be cases for which the lengths of the 95% confidence intervals are larger than 0.15, and neither of the two desired precision levels can be achieved. This is the main reason why the number of runs need to be increased to at least 100 runs or more for these cases.

The convergence of the Avg. 99th PC values obtained using 100, 200, and 300 runs, as the number of iterations increases, can be seen in Figure 17, which shows the Avg. 99th PC values obtained for the case shown in Figure 3. The plot shows the Avg. 99th PC values obtained from the four tested scenarios. For each scenario, four sets of either 100, 200, or 300 random seeds were used, and the same sets of random seeds were used at 300 runs for both 50,000 and 100,000 iterations. The plot shows that as the number of total iterations increases, the Avg. 99th PC values obtained become more stable with respect to changing the set of random seeds. The lengths of the 95% confidence intervals for the Avg. 99th PC values from Scenario 2 are less than 0.15, and the lengths of the 95% confidence intervals for the Avg. 99th PC values from Scenarios 3 and 4 are less than 0.10. There is a clear trend of the Avg. 99th PC values converging to a stable value that represents the theoretical 99th PC value of the distribution of PC values for this case. The length of the 95% confidence interval for the theoretical 99th PC value obtained from a 15,000,000 sample is 0.10.

Figure 18 shows the convergence of the Avg. 99th PC values, as the number of iterations increases, for the case shown in Figure 4. The plot shows that as the number of total iterations increases, the Avg. 99th PC values obtained become more stable with respect to changing the set of random seeds. The lengths of the 95% confidence intervals for the Avg. 99th PC values from Scenarios 2 and 3 are less than 0.06, and the lengths of the 95% confidence intervals for the Avg. 99th PC values from Scenario 4 are less than 0.04. There is a clear trend of the Avg. 99th PC values converging to a stable value that represents the theoretical 99th PC value of the distribution of PC values for this case. The length of the 95% confidence interval for the theoretical 99th PC value obtained from a 15,000,000 sample is 0.06. However, in all four scenarios, for which between 5,000,000 to 30,000,000 iterations were used, the compensability of this case is hard to decide, since the Avg. 99th PC values bounce around the 50.0% compensability threshold.

Figure 19 shows the convergence of the Avg. 99th PC values, as the number of iterations increases, for the case shown in Figure 5. The plot shows that as the number of total iterations increases, the Avg. 99th PC values obtained become more stable with respect to changing the set of random seeds. The lengths of the 95% confidence intervals for the Avg. 99th PC values from Scenario 3 are less than 0.10, and the lengths of the 95% confidence intervals for the Avg. 99th PC values from Scenario 4 are less than 0.07. There is a clear trend of the Avg. 99th PC values converging to a stable value that represents the theoretical 99th PC value of the distribution of PC values for this case. The length of the 95% confidence interval for the theoretical 99th PC value obtained from a 20,000,000 sample is 0.09. Using the first two scenarios, the compensability of this case is hard to decide, since the Avg. 99th PC values bounce around the 50.0% compensability threshold. As the total of number of runs and iterations are increased in the last two scenarios, the Avg. 99th PC values become more stable, but it is still hard to reach a decision about the compensability. These last two scenarios suggest that for situations where the compensability is hard to decide, increasing the number of iterations such that the length of 95% confidence interval is less than 0.07, will make a compensability decision easier to reach since the Avg. 99th PC values becomes much more stable.

Plots showing the convergence of the Avg. 99th PC values from 100 or more runs, as the sample size increases, are shown in Appendix B for 15 additional cases with Avg. 99th PC values between 49.00% and 49.99%. The 95% confidence intervals for the theoretical 99th PC value for most of these cases were obtained at 5,000,000 iterations; the number of iterations was increased to 10,000,000 or 15,000,000 iterations for a few cases for which the Avg. 99th PC value was either close to the 50.0% compensation threshold or for which the length of the 95% confidence interval for the theoretical 99th PC value was too large at 5,000,000 iterations.

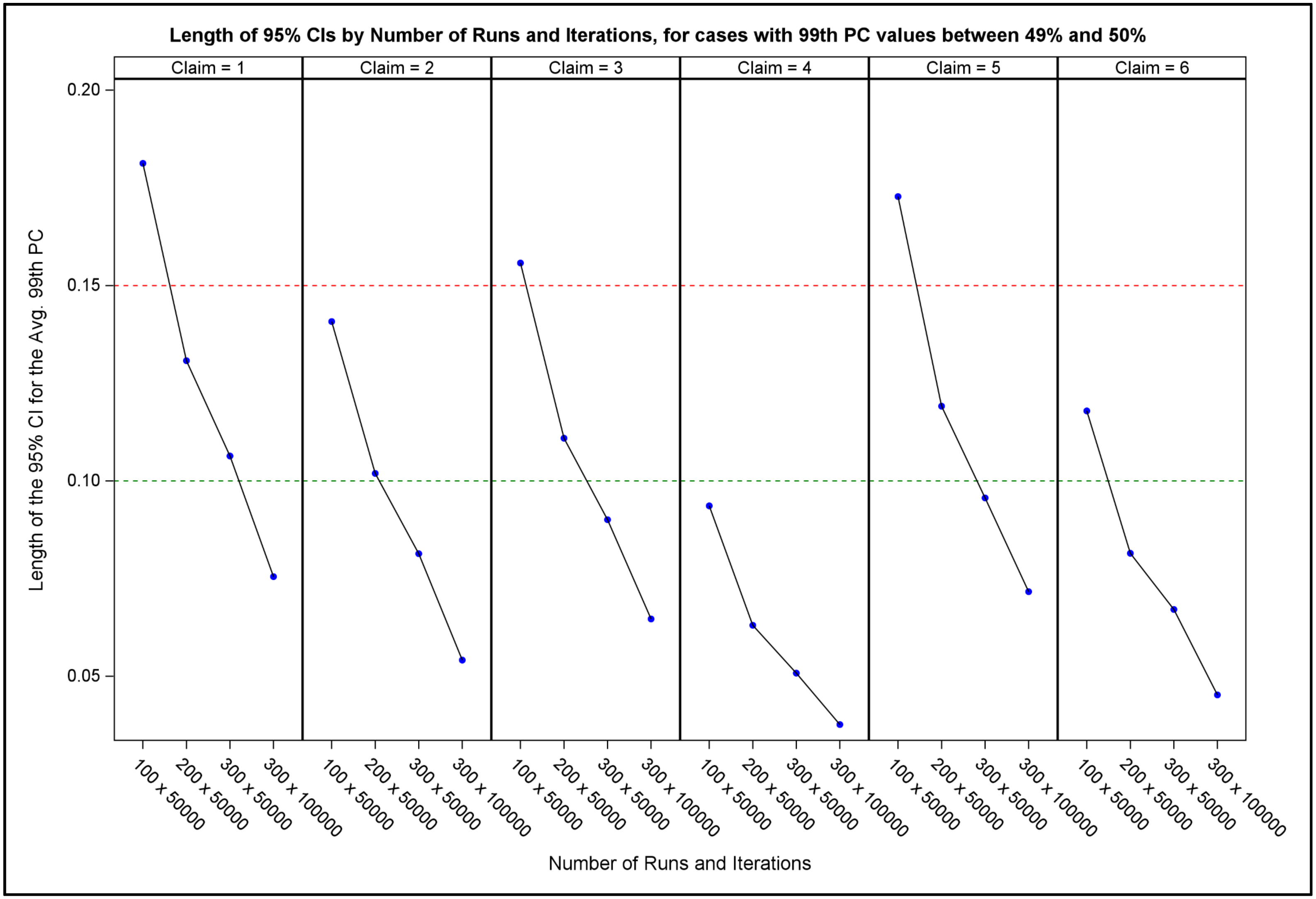


Figure 7. Lengths of the 95% CIs from four scenarios with different number of runs and iterations (part 1).

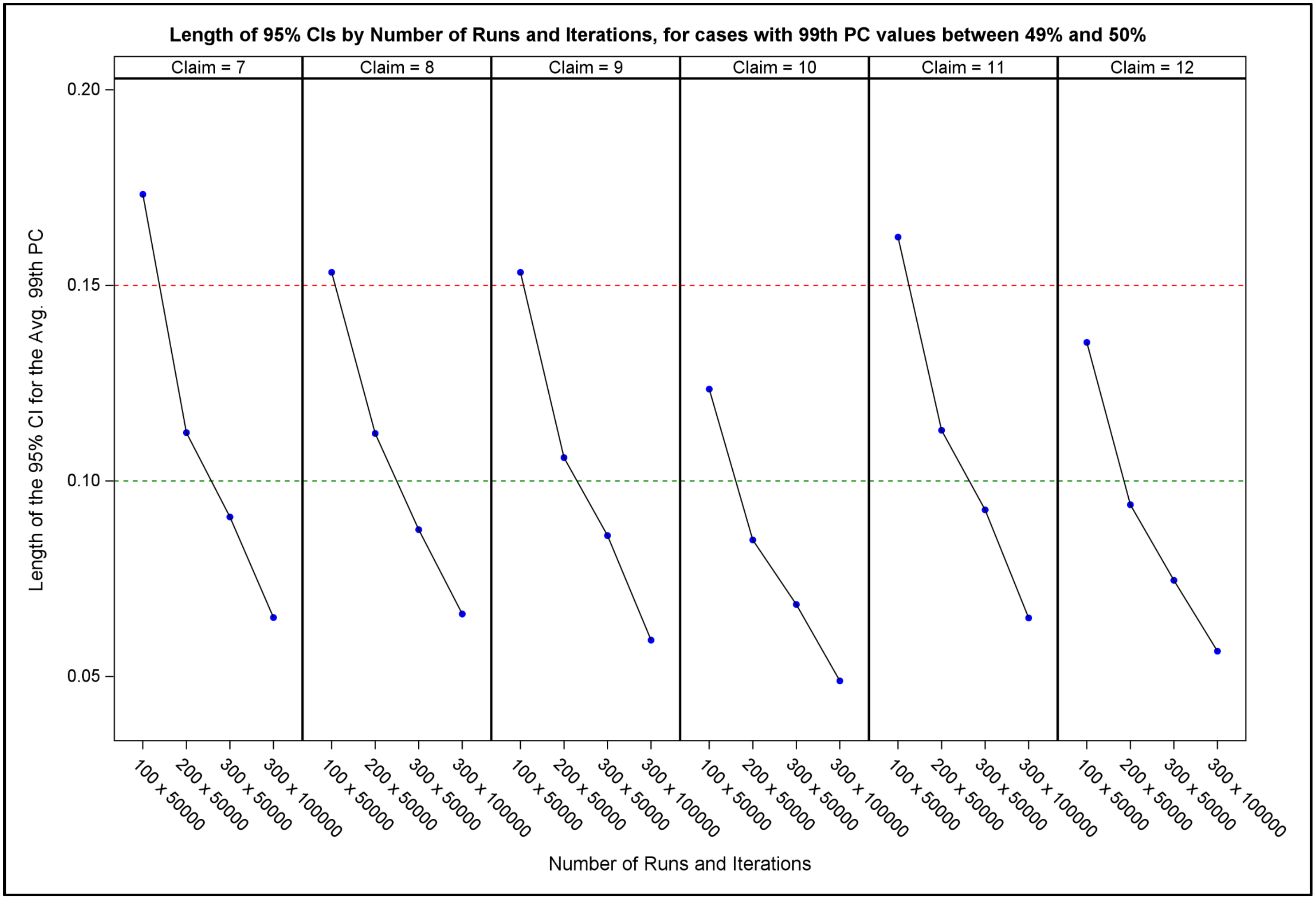


Figure 8. Lengths of the 95% CIs from four scenarios with different number of runs and iterations (part 2).

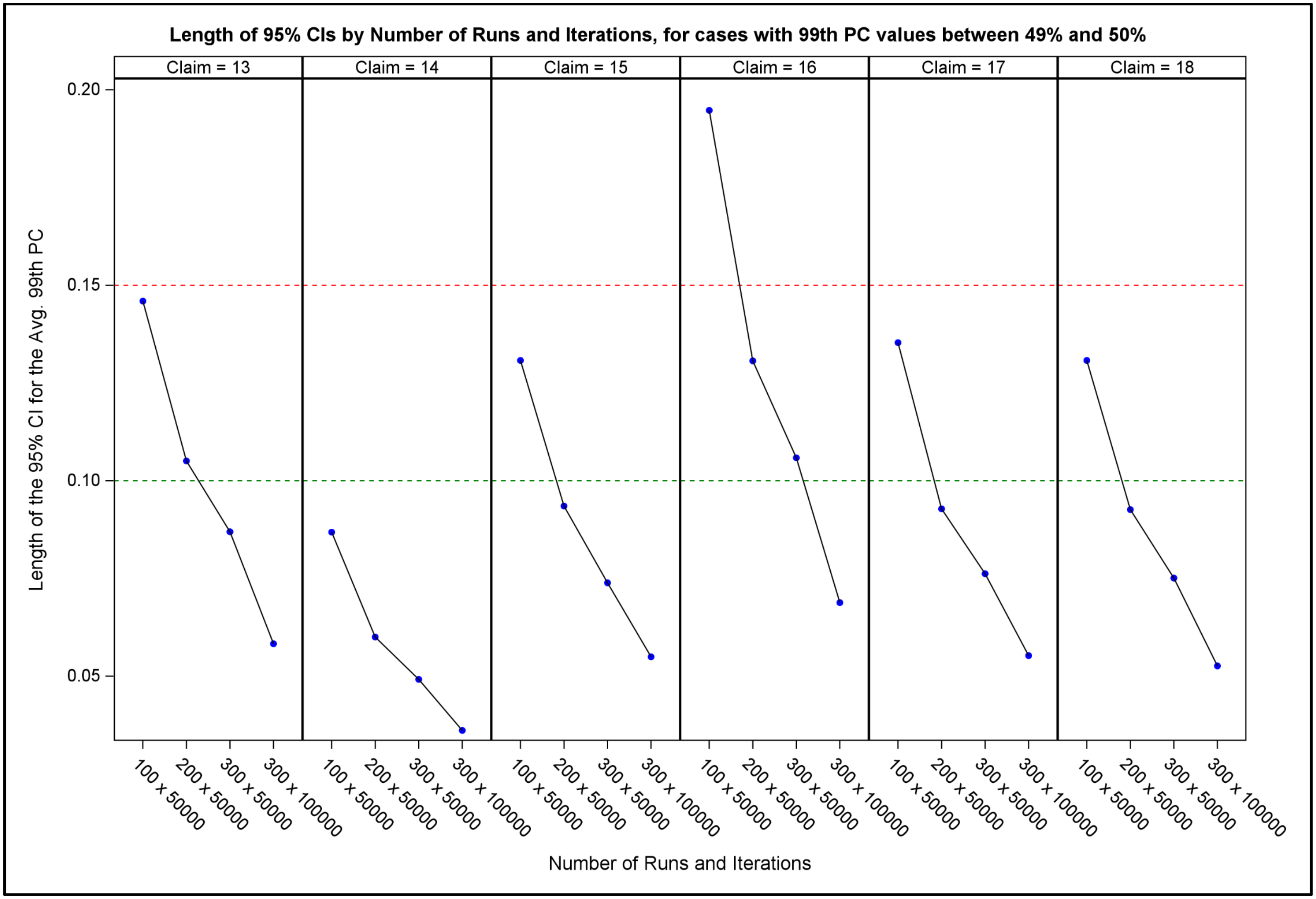


Figure 9: Lengths of the 95% CIs from four scenarios with different number of runs and iterations (part 3).

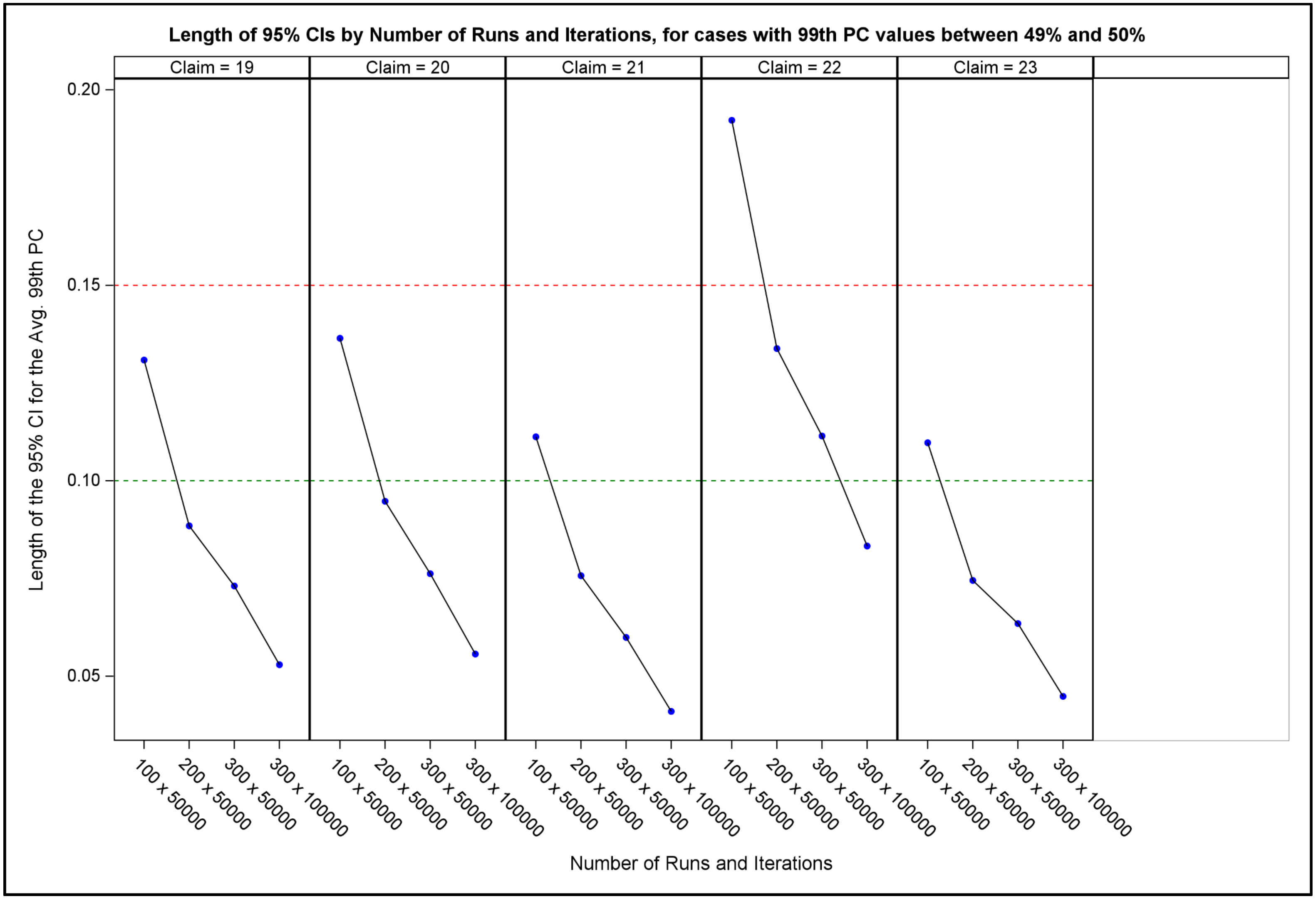


Figure 10. Lengths of the 95% CIs from four scenarios with different number of runs and iterations (part 4).

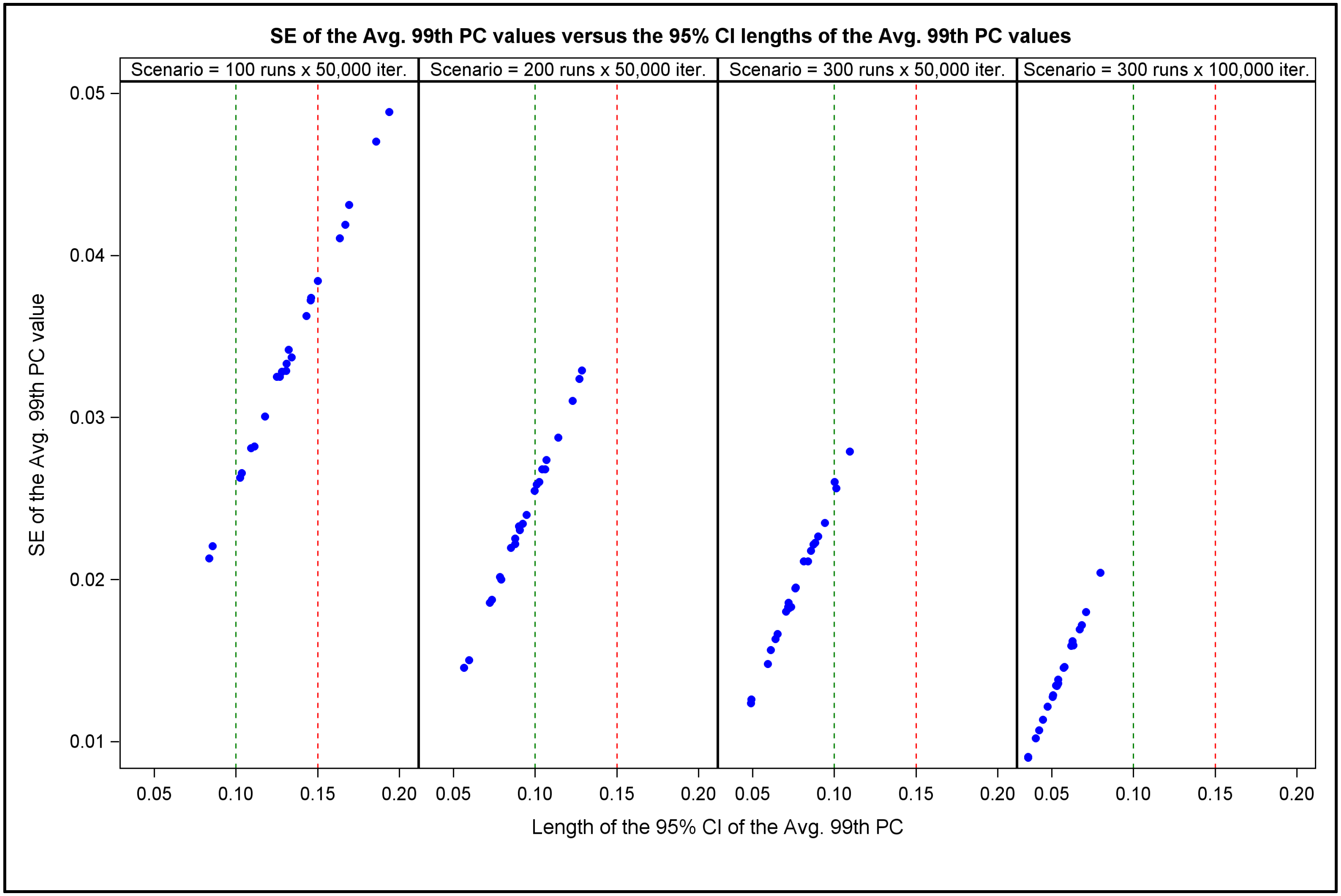


Figure 11. Standard errors (SE) versus the length of the 95% CIs from the four scenarios with different number of runs and iterations.

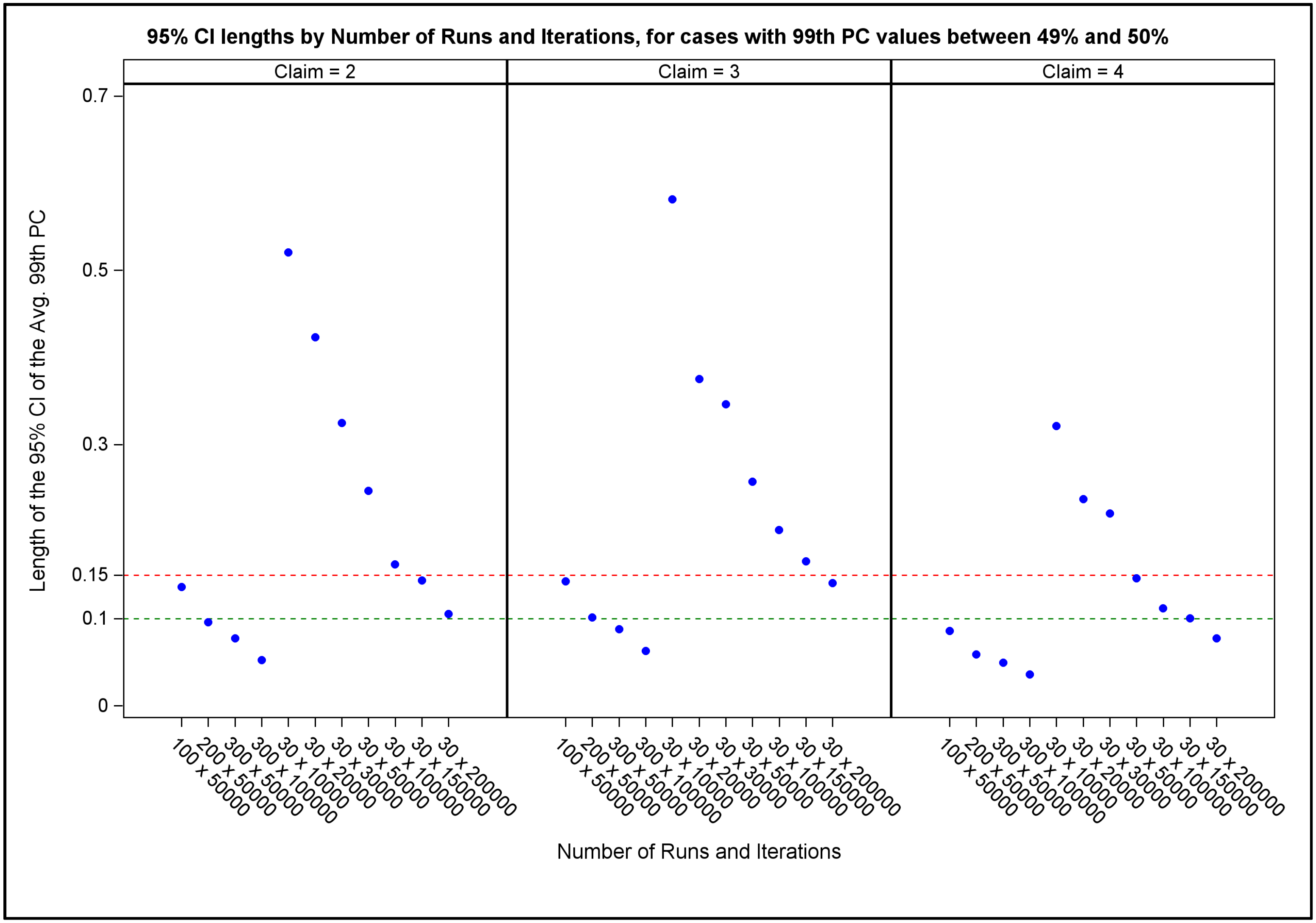


Figure 12. Lengths of the 95% CIs of the Avg. 99th PC from 11 scenarios with different number of runs and iterations (part 1).

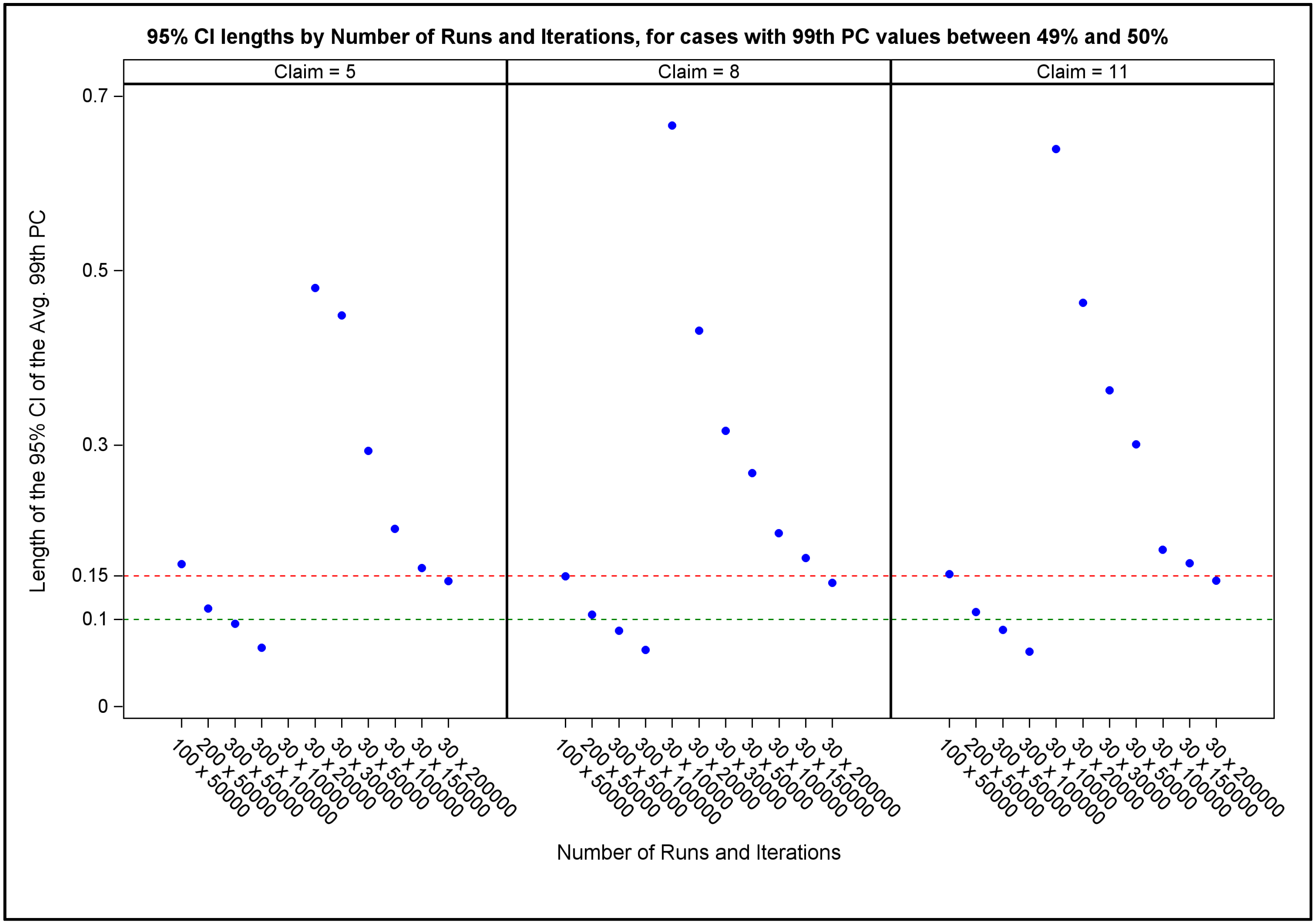


Figure 13. Lengths of the 95% CIs of the Avg. 99th PC from 11 scenarios with different number of runs and iterations (part 2).

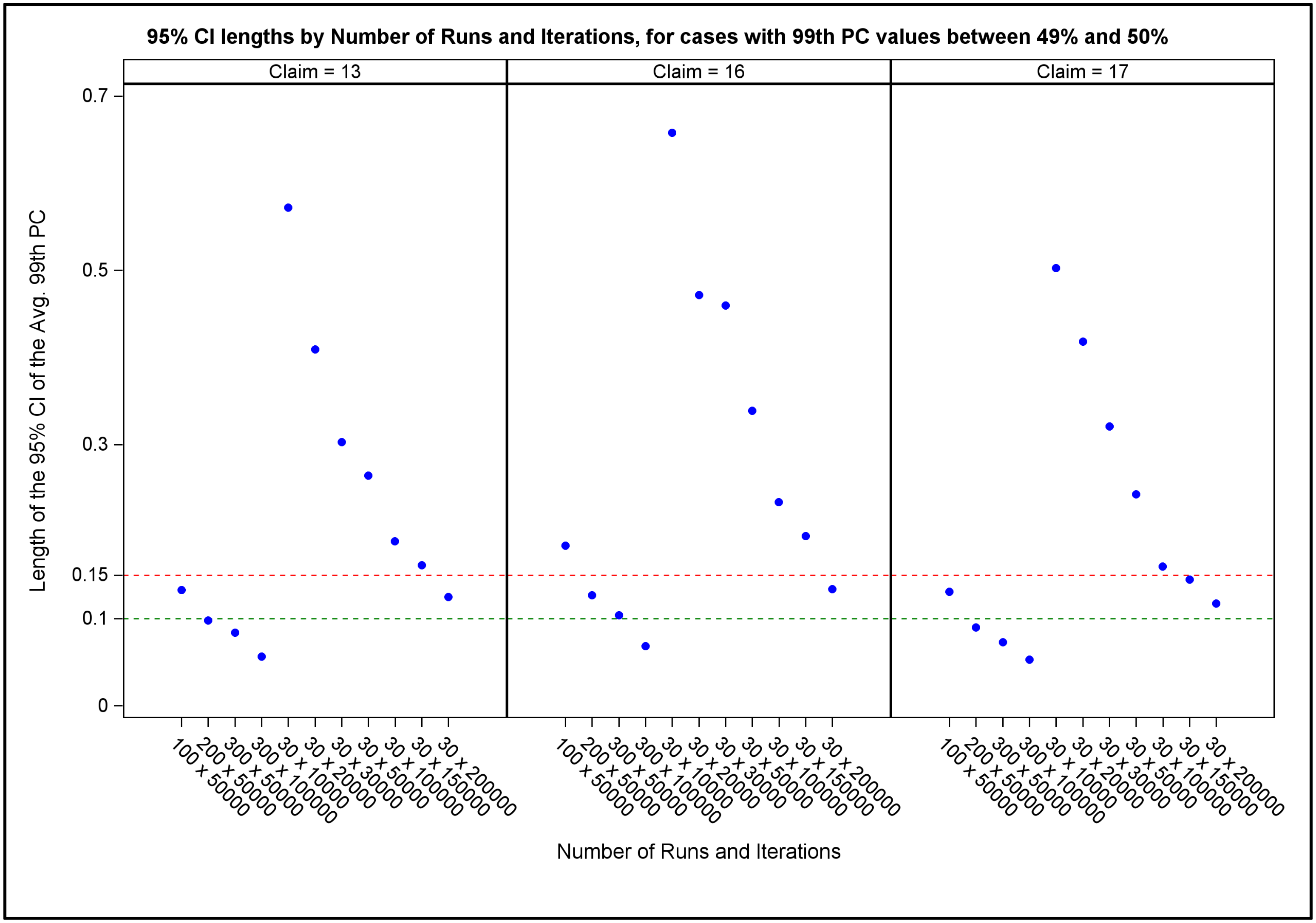


Figure 14. Lengths of the 95% CIs of the Avg. 99th PC from 11 scenarios with different number of runs and iterations (part 3).

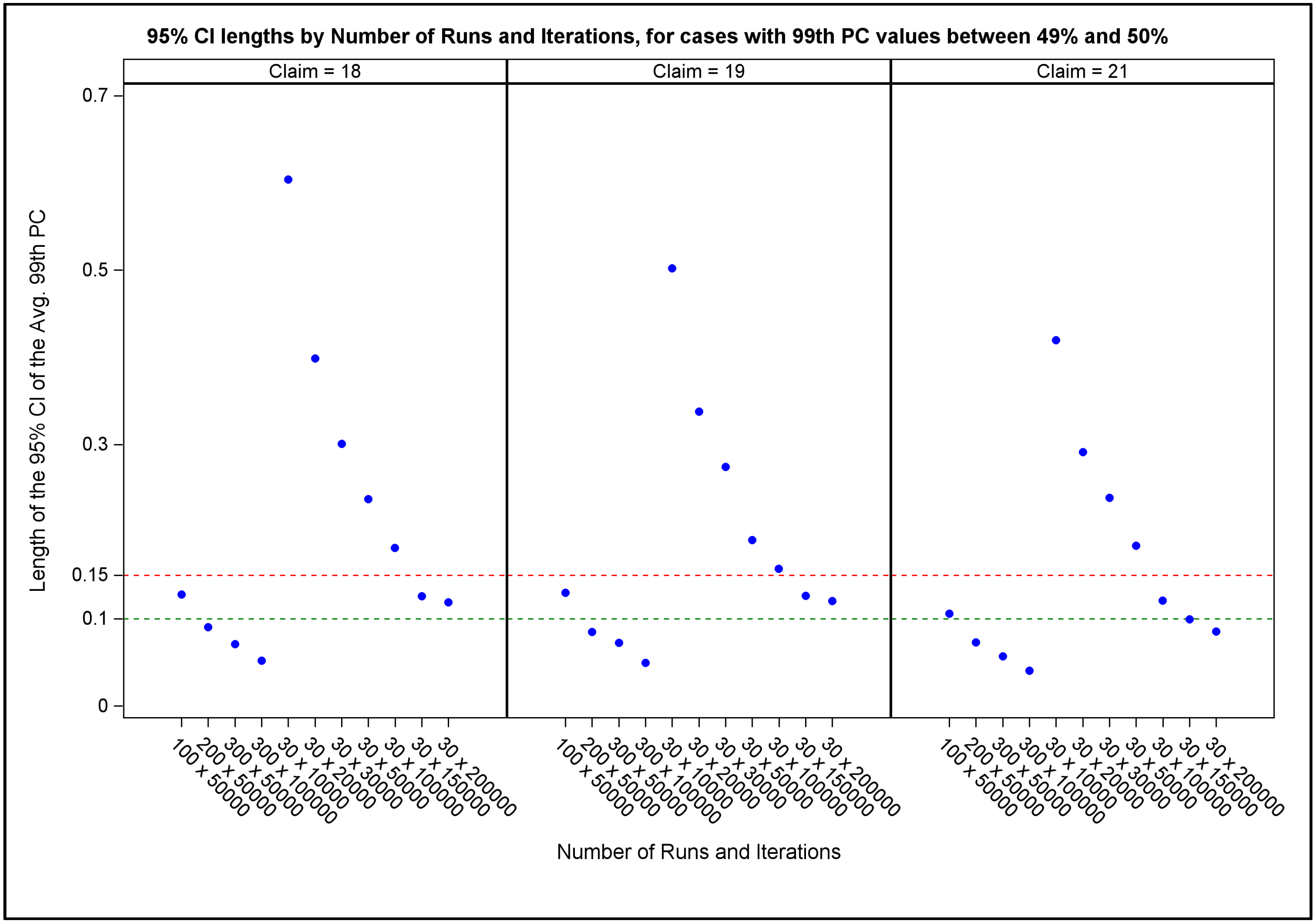


Figure 15. Lengths of the 95% CIs of the Avg. 99^{th} PC from 11 scenarios with different number of runs and iterations (part 4).

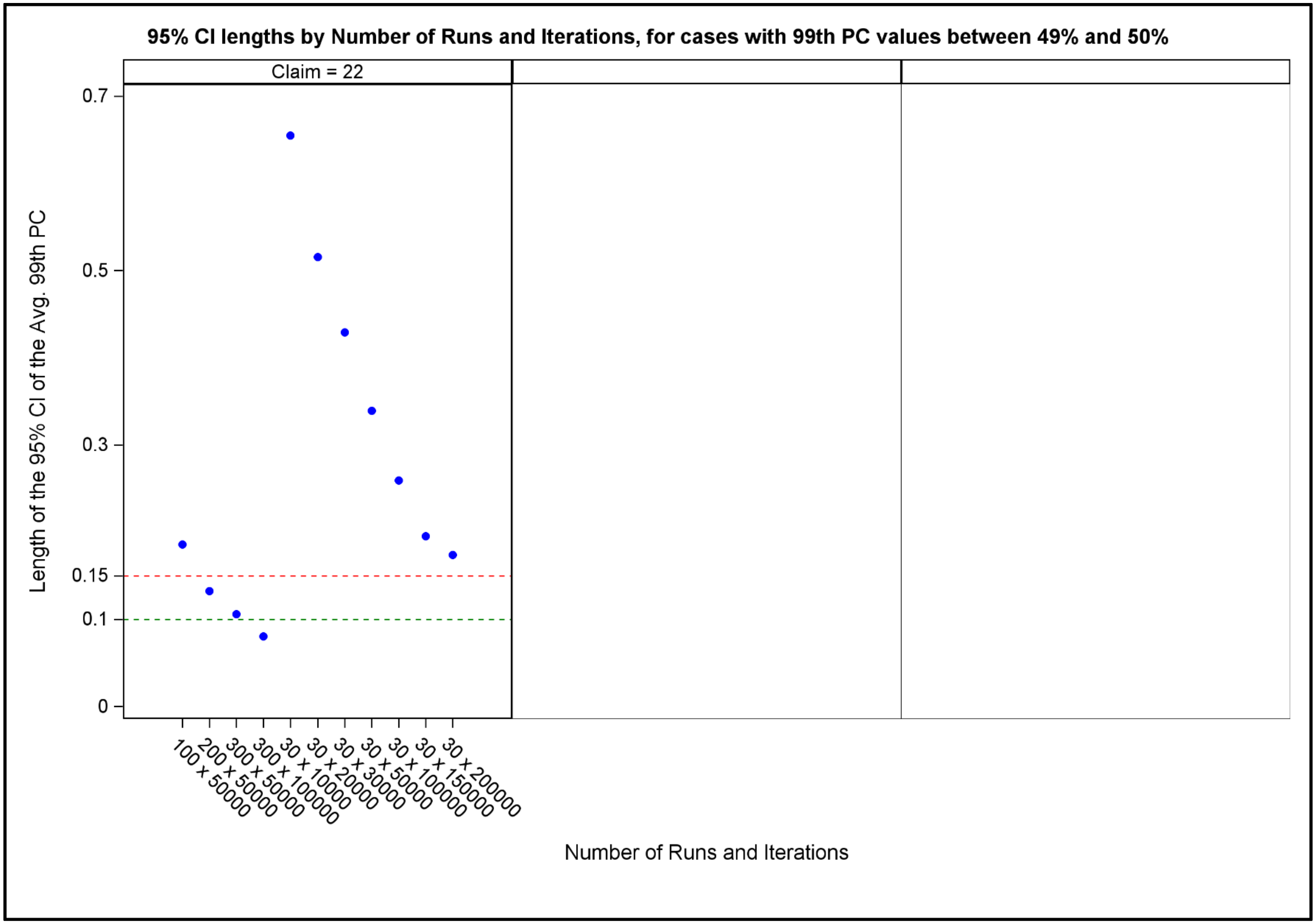


Figure 16. Lengths of the 95% CIs of the Avg. 99^{th} PC from 11 scenarios with different number of runs and iterations (part 5).

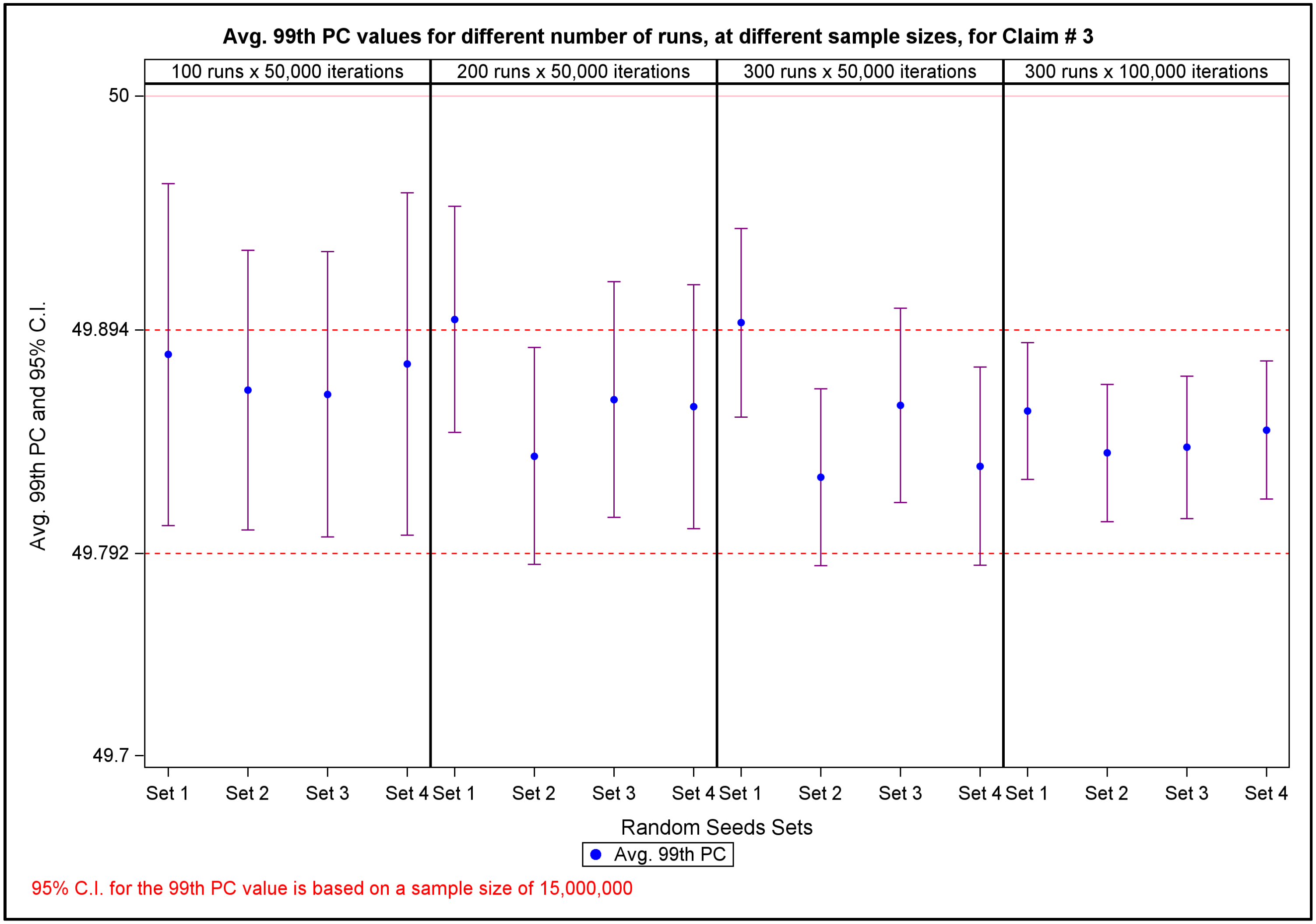


Figure 17. Avg. 99th PC values from 100, 200, and 300 runs, and their 95% CIs, and the confidence limits for the theoretical 99th PC value.

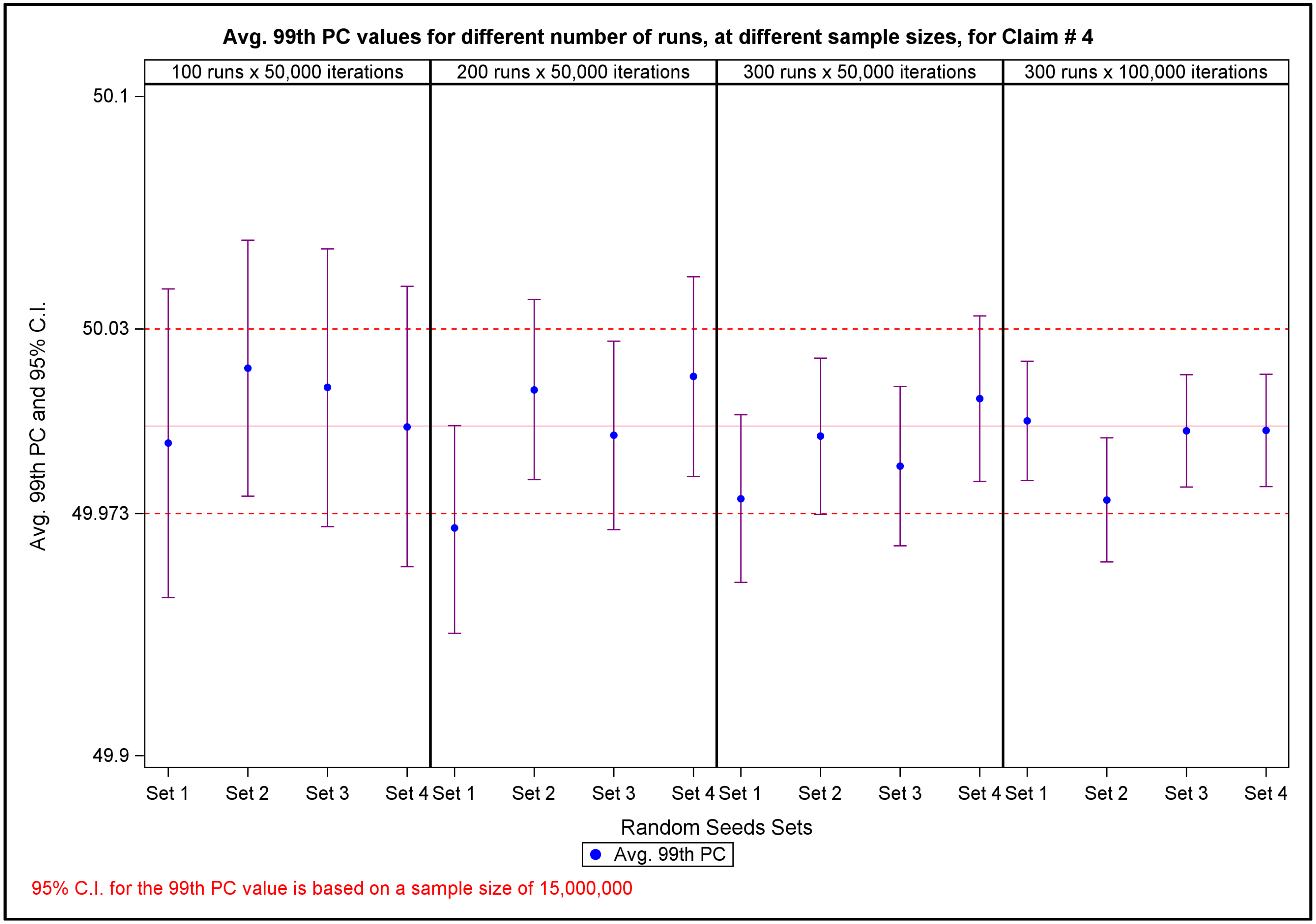


Figure 18. Avg. 99th PC values from 100, 200, and 300 runs, and their 95% CIs, and the confidence limits for the theoretical 99th PC value.

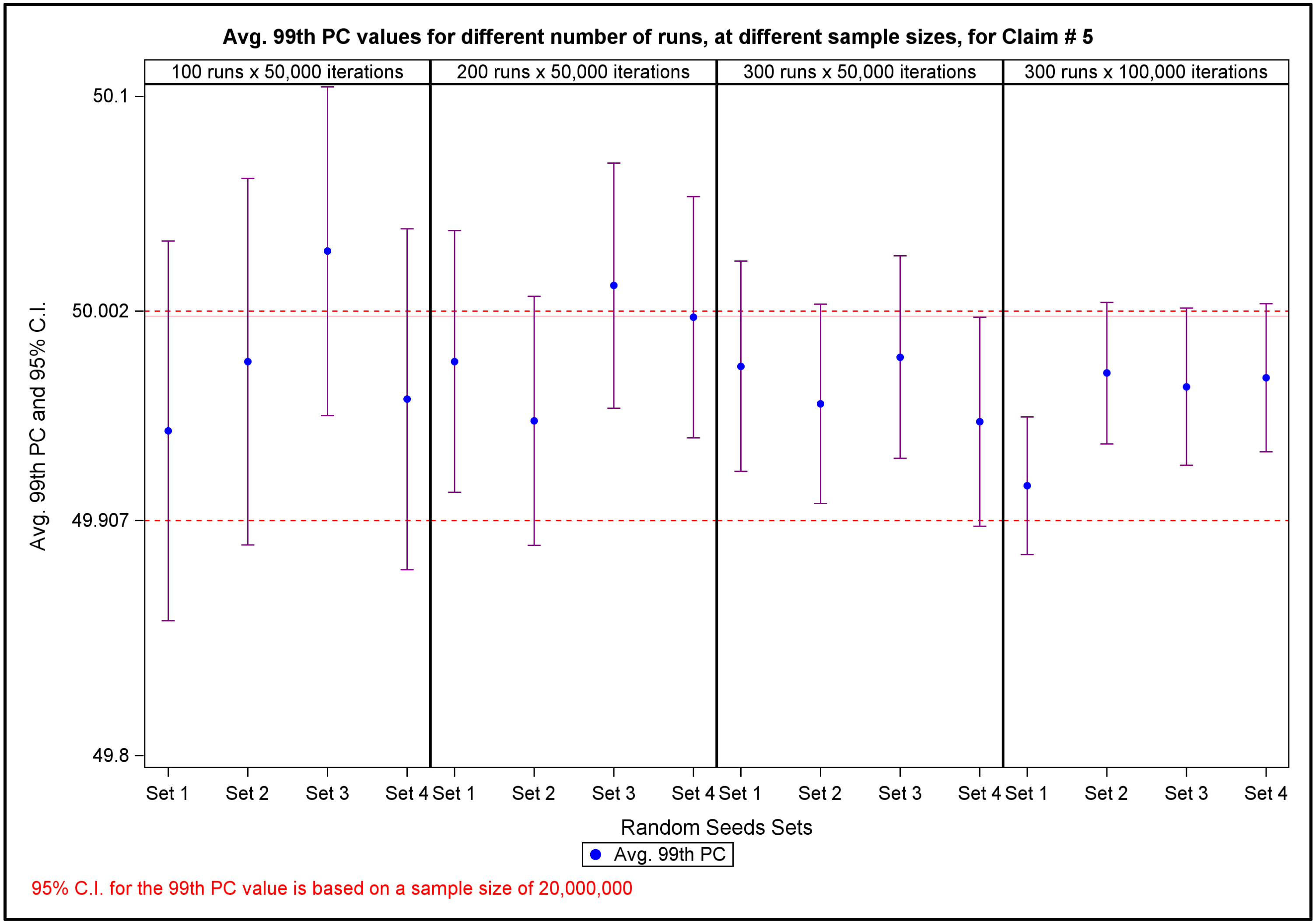


Figure 19. Avg. 99th PC values from 100, 200, and 300 runs, and their 95% CIs, and the confidence limits for the theoretical 99th PC value.

5.0 IREP PREDICTIVE TOOL TO MINIMIZE THE NUMBER OF ITERATIONS

In the previous section, it was shown that for the 23 tested cases with Avg. 99th PC values between 49.00% and 50.01%, the lengths of the 95% confidence intervals are smaller than 0.15 for all cases in Scenarios 2 and 3, and less than 0.10 for all cases in Scenario 4. However, in each of these scenarios, the lengths of the 95% confidence interval are sometimes much smaller than the desired precision levels, which suggests that the actual number of iterations needed might be even smaller than the fixed number of iterations set up in these four scenarios, depending on factors specific to each case. Reducing the number of iterations for a case, while still being able to achieve the desired precision for the 95% confidence interval of the Avg. 99th PC values, will help with reducing the IREP running times associated with completing each case, which for Scenario 2 was around 2 hours for the case with the largest number of exposures, and for Scenario 3 was around 3 hours for the case with the largest number of exposures.

To minimize the number of iterations needed to achieve the desired precision for the length of the 95% confidence interval of the Avg. 99th PC values, a predictive tool was built into the Research version of IREP Enterprise Edition. This IREP predictive tool uses an observed decreasing power function relating the length of the 95% confidence interval of the Avg. 99th PC values to the number of iterations, that was fitted based on multiple trials with different numbers of iterations for the tested cases. The function has an estimated exponent of -0.5. To use the IREP predictive tool for a single cancer case, the following two steps are needed:

1) Run IREP Enterprise Edition using 300 runs with a small number of iterations (e.g., 1,000 iterations). If the precision level is not already met, the built-in tool will display a recommended number of iterations that will help to reach the desired precision level.
2) Run IREP Enterprise Edition using 300 runs and the recommended number of iterations (or larger) to obtain Avg. 99th PC values with the desired precision for the 95% confidence interval.

Each of the 23 cases previously selected were also tested using the IREP Predictive Tool, to determine the minimum number of iterations needed to reach the 0.15 and the 0.10 precision criteria for each case.

Figure 20 shows a plot with the IREP EE running times obtained for the two desired precision levels using the number of iterations determined according to the IREP Predictive Tool. The plot shows the IREP EE times (with 300 runs) versus the number of exposures for all 23 cases with Avg. 99th PC values between 49.00% and 50.01%. The largest IREP running time observed for the 23 cases when the precision level was 0.15 is less than 1 hour, and the largest IREP running time observed when the precision level was 0.10 is around 1.75 hours.

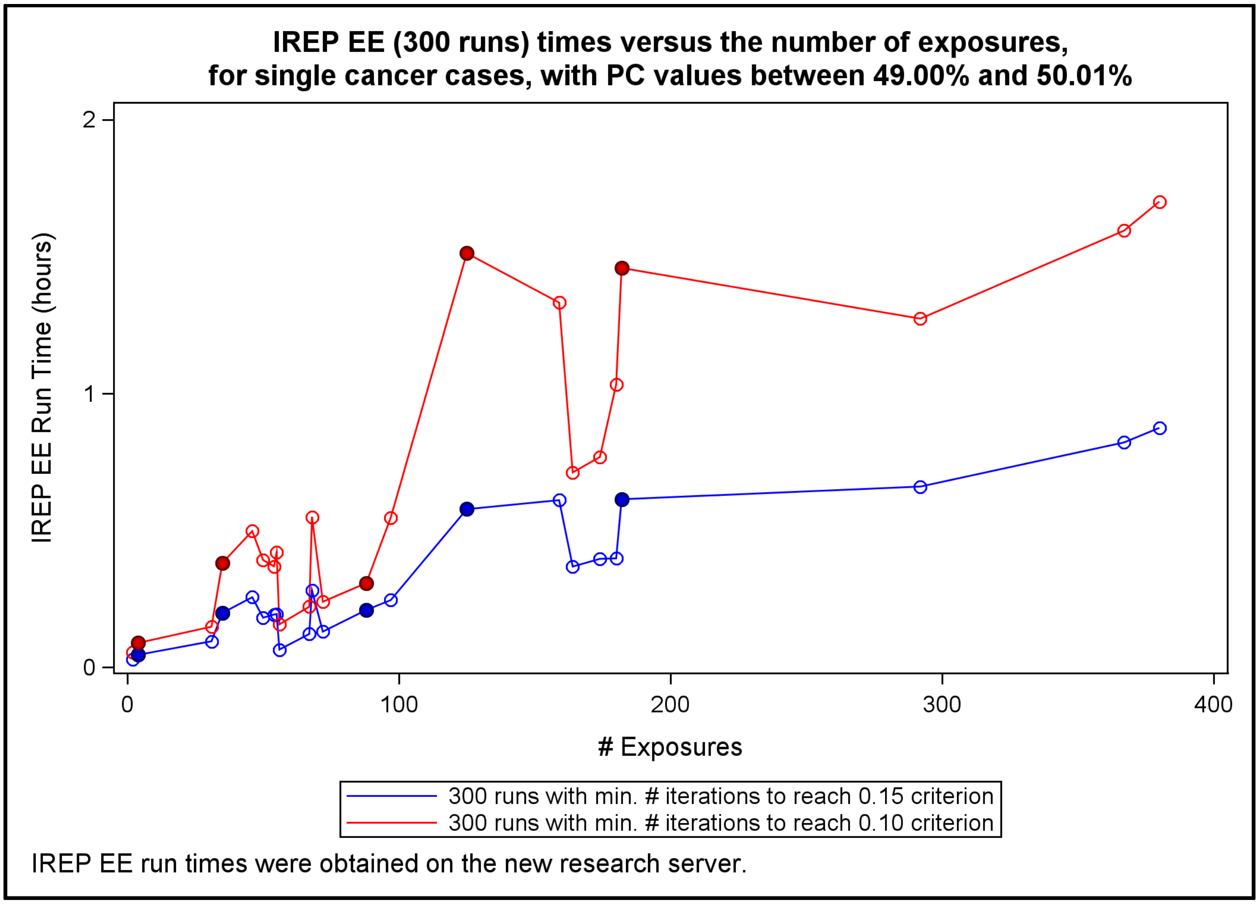


Figure 20. IREP EE running times for the 23 single cancer cases with Avg. 99th PC values between 49.00% and 50.01%.

Figure 21 shows the minimum number of iterations needed for each case to reach the two desired precision levels. For the 0.15 precision level, the number of iterations needed for the 23 cases varies between 5,000 and 35,000 iterations, which is a much smaller number of iterations than the number of iterations used in Scenario 2, in which a fixed number of 50,000 iterations were used for all cases, and which was enough to reach the 0.15 precision level for all cases. For the 0.10 precision level, the number of iterations needed for the 23 cases varies between 15,000 and 60,000 iterations, which also represent a much smaller number of iterations than the number of iterations used in Scenario 4, in which a fixed number of 100,000 iterations were used for all cases, and which was enough to reach the 0.10 precision level.

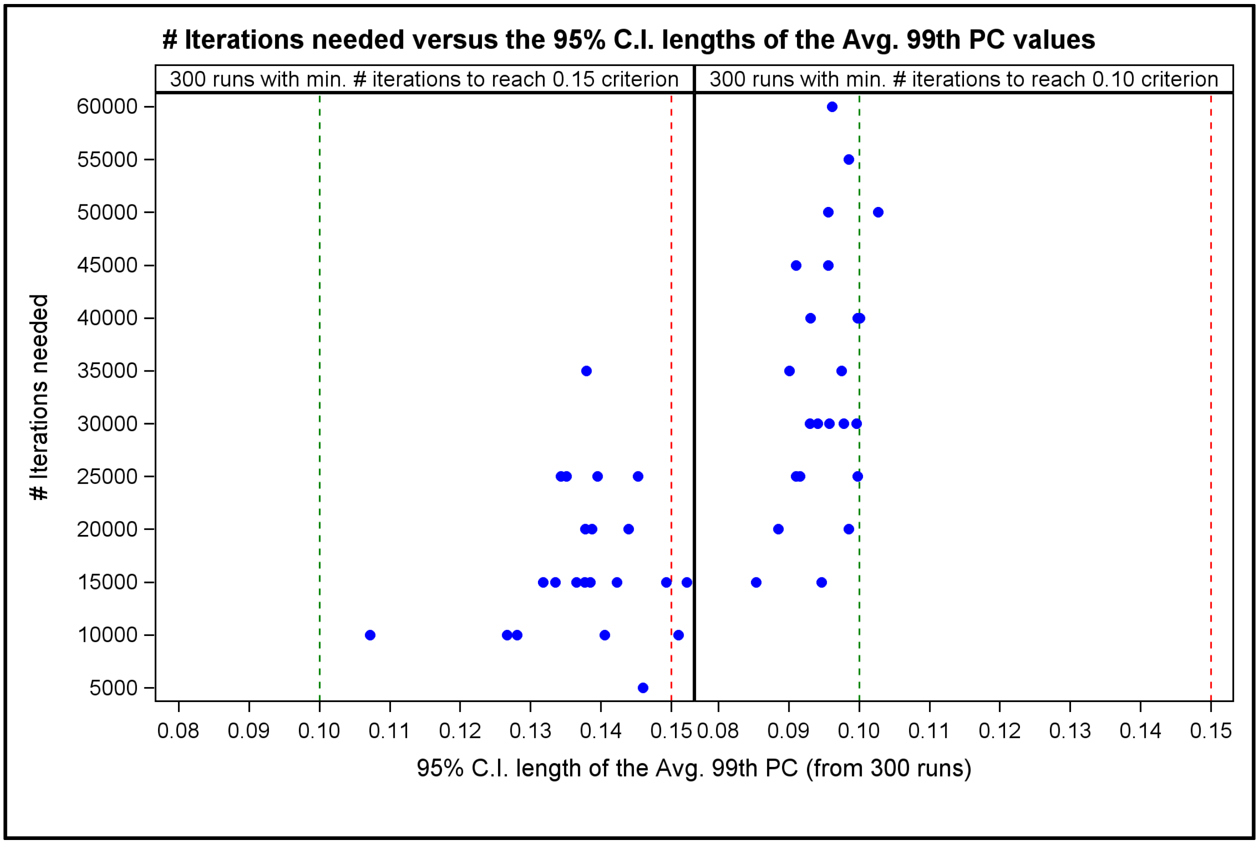


Figure 21. Minimum number of iterations needed for the 23 single cancer cases with 99th PC values between 49.00% and 50.01%, in order to reach the desired precision levels.

In Figure 11, it was determined that for the tested cases, the lowest length of the 95% confidence intervals around the Avg. 99th PC values was 0.05 using 300 runs and 50,000 iterations, and the lowest length of the 95% confidence intervals was 0.03 using 300 runs and 100,000 iterations. Figure 21 shows that somewhere between 15,000 and 60,000 iterations are needed at 300 runs to achieve a 95% confidence interval of 0.1 or less for all the tested cases. This indicates that running a case with 300 runs and a flexible number of iterations per run, somewhere between 15,000 and 60,000 iterations will be needed to obtain a 95% confidence interval of 0.1, and that 95% confidence intervals between 0.05 and 0.1 can be achieved for most of the cases if completed with 300 runs and a flexible number of iterations somewhere between 50,000 and 100,000. This was indeed confirmed using the IREP predictive tool that can also determine the number of iterations per run, at 300 runs, that is needed to achieve the lower precision levels of 0.07 or 0.05.

Based on previous findings, running a case with 300 runs and somewhere between 50,000 and 100,000 iterations per run, should result in the length of the 95% confidence interval around the Avg. 99th PC values between 0.05 and 0.10, and this is the lowest precision that can achieved for most of the cases. So, the precisions around the Avg. 99th PC values obtained in IREP EE can be estimated somewhere between 0.05 and 0.10 if enough iterations are used. This indicates there is a limit with which the precision around the Avg. 99th PC values in IREP EE can be estimated. A precision lower than 0.05 will be either difficult or even impossible to reach for most cases, since it will require an extensive running time in IREP, causing the calculation to become prohibitive.

The previous 23 cases tested were all single cancer cases. The procedure of using the IREP Predictive Tool for cases with multiple cancers is addressed below.

To use the IREP predictive tool for a case with multiple cancers, the following three steps are needed:

1) Run IREP Enterprise Edition using 300 runs with a small number of iterations (e.g., 1,000 iterations) for each cancer. If the precision level is not already met, the built-in tool will display a recommended number of iterations to reach the desired precision level.
2) From the recommended numbers of iterations for all cancers, determine the largest recommended number of iterations.
3) Run IREP Enterprise Edition using 300 runs and the largest recommended number of iterations to obtain Avg. 99th PC values with the desired or better precision for all cancers.

The use of the IREP Predictive Tool for cases with multiple cancers is a conservative procedure which guarantees that the Combined Avg. 99th PC value (obtained using the Multiple Cancer formula in IREP) will have the desired precision for the 95% confidence interval around the Combined Avg. 99th PC value. If the Combined Avg. 99th PC value for all the cancers is lower than the desired precision level, and the number of iterations required for all the cancers is too large, then the number of iterations can be lowered through an optimization process, such that the Combined Avg. 99th PC value for all the cancers will have the desired precision level. However, to minimize the bias that was observed at the 99th PC values between using the default and the alternative percentile definitions (Stancescu et al. 2021), it is recommended to use a minimum of 20,000 iterations for both single and multiple cancer cancers, even when the IREP Predictive Tool might suggest a lower number of iterations.

6.0 VALIDATION OF 99TH PC VALUES FROM IREP AND IREP EE VERSIONS

As a validation exercise, this section estimates and compares the 99th PC values obtained from the IREP version using one run with a large number of iterations versus the Avg. 99th PC values obtained from the IREP Enterprise Edition using multiple runs with smaller numbers of iterations.

The equivalence of the 99th PC values obtained in the IREP and the Avg. 99th PC values obtained from the IREP EE with the current settings in version v.5.9 (30 runs with 10,000 iterations) is illustrated in Figures 22 and 23. The total number of iterations used in both versions is the same, with IREP using a single run with 300,000 iterations, and the IREP EE using 30 runs with 10,000 iterations. The plots show the 99th PC values obtained from the two IREP versions for six different cases (all six cases have Avg. 99th PC values on record between 49% and 52%), for 30 random seeds used in IREP, and for 30 sets of 30 random seeds used in IREP EE. The plots also show the 95% confidence intervals around each of the 99th PC values from IREP, and around the Avg. 99th PC values from IREP EE. An examination of the plots reveals that the lengths of the 95% confidence intervals are very similar in the two IREP versions.

The equivalence between the 99th PC values obtained in the IREP version and the Avg. 99th PC values obtained from the IREP EE version is preserved when the total number of iterations increases significantly (Figures 24 through 29). Each figure shows the 99th PC values obtained in the IREP version and the Avg. 99th PC values obtained from the IREP EE version for a different case, and for each case three different scenarios with a different number of iterations were used:

- 5,000,000 iterations (single IREP run) vs. 100 runs with 50,000 iterations (in IREP EE)
- 10,000,000 iterations (single IREP run) vs. 200 runs with 50,000 iterations (in IREP EE)
- 15,000,000 iterations (single IREP run) vs. 300 runs with 50,000 iterations (in IREP EE)

Each of the three scenarios in these plots shows the 99th PC values obtained from the two IREP versions for 20 random seeds used in IREP, and for 20 sets with either 100, 200, or 300 random seeds used in IREP EE. The red lines in each plot represent the average between the 99th PC value and the Avg. 99th PC value obtained using 15,000,000 iterations, which are the most accurate estimates for the theoretical 99th PC value of the distribution of PC values for each case.

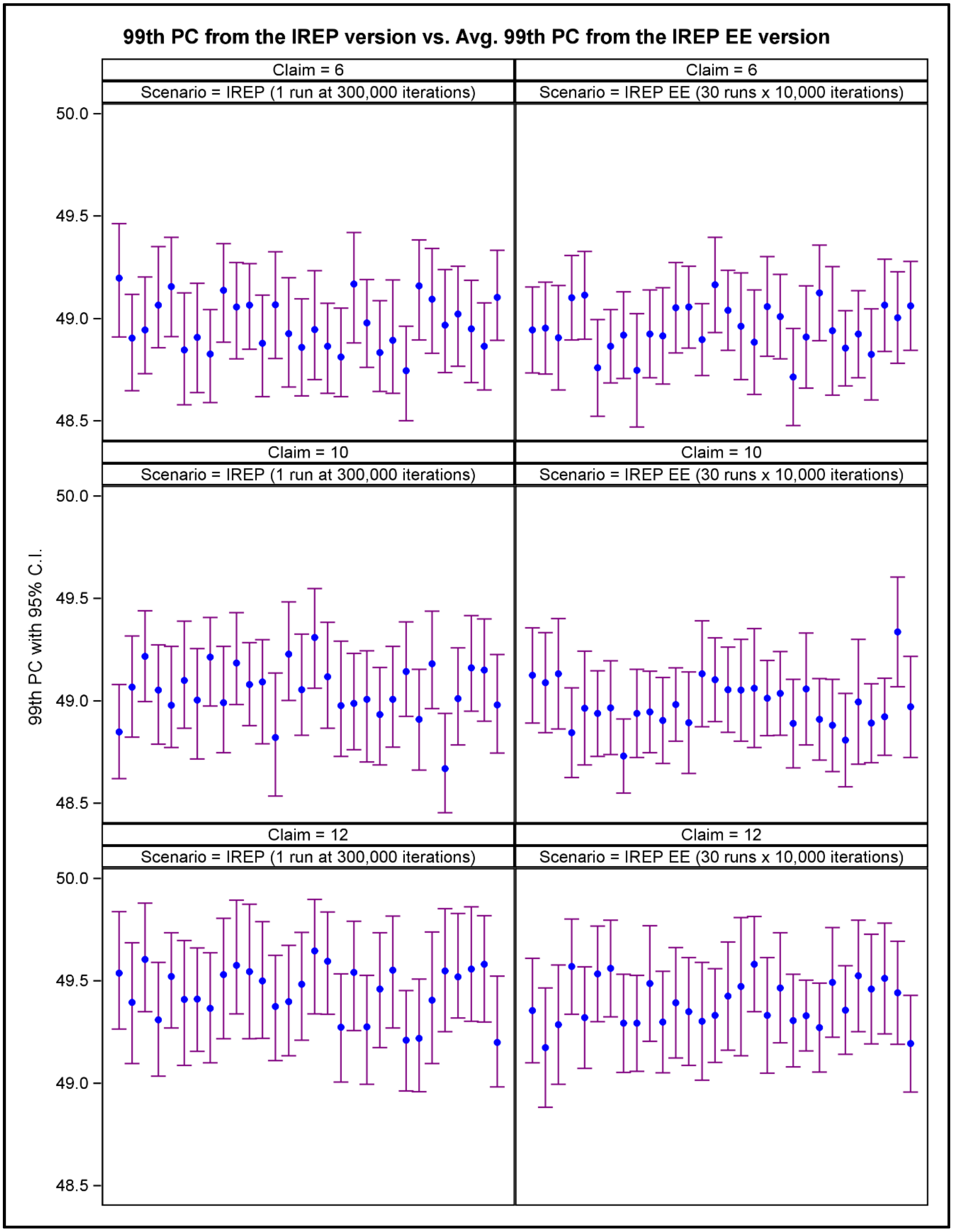


Figure 22. The 99th PC values obtained from the IREP version, versus Avg. 99th PC values from the IREP EE version, obtained at 300,000 iterations, for three different cases.

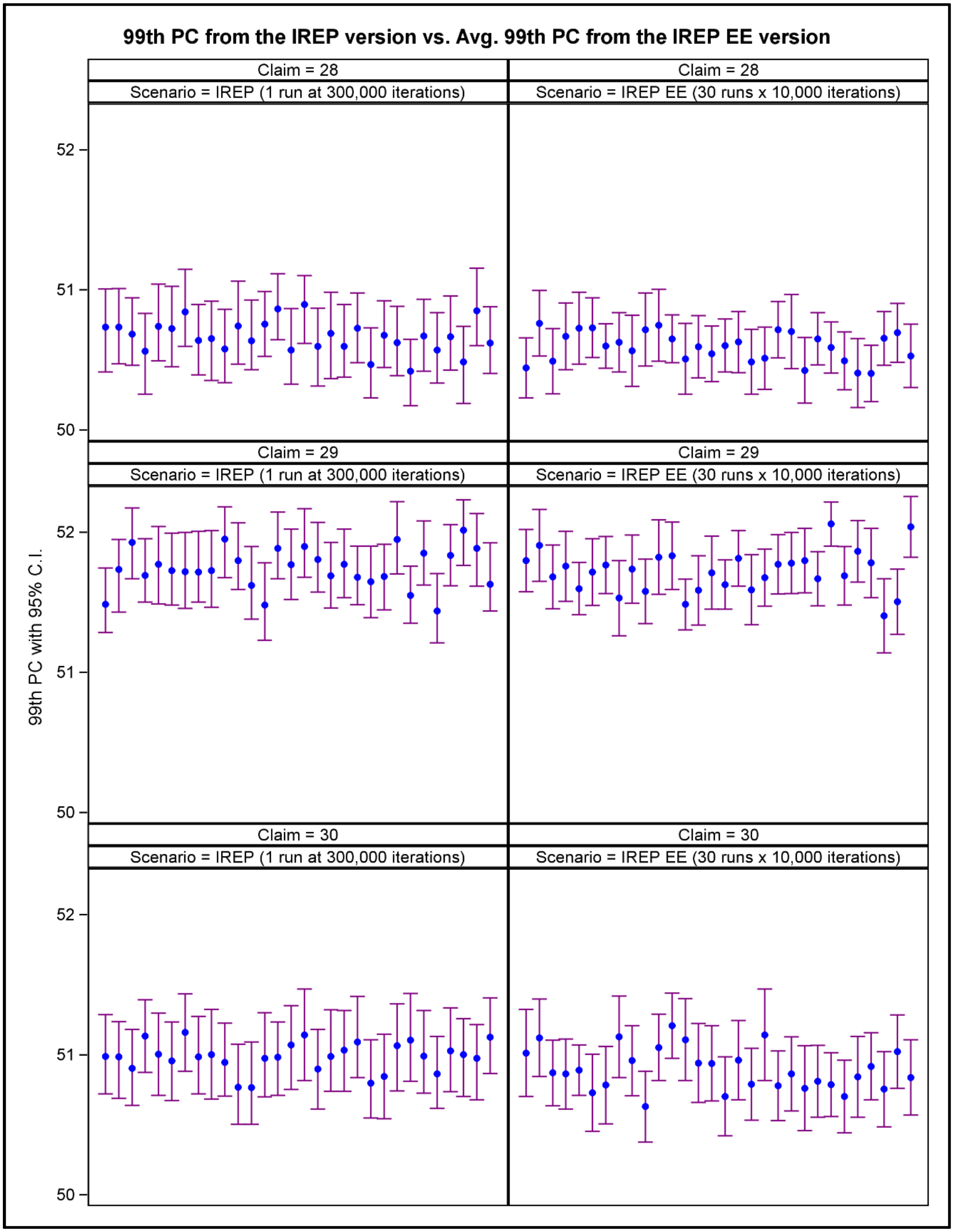


Figure 23. The 99th PC values obtained from the IREP version, versus Avg. 99th PC values from the IREP EE version, obtained at 300,000 iterations, for three different cases.

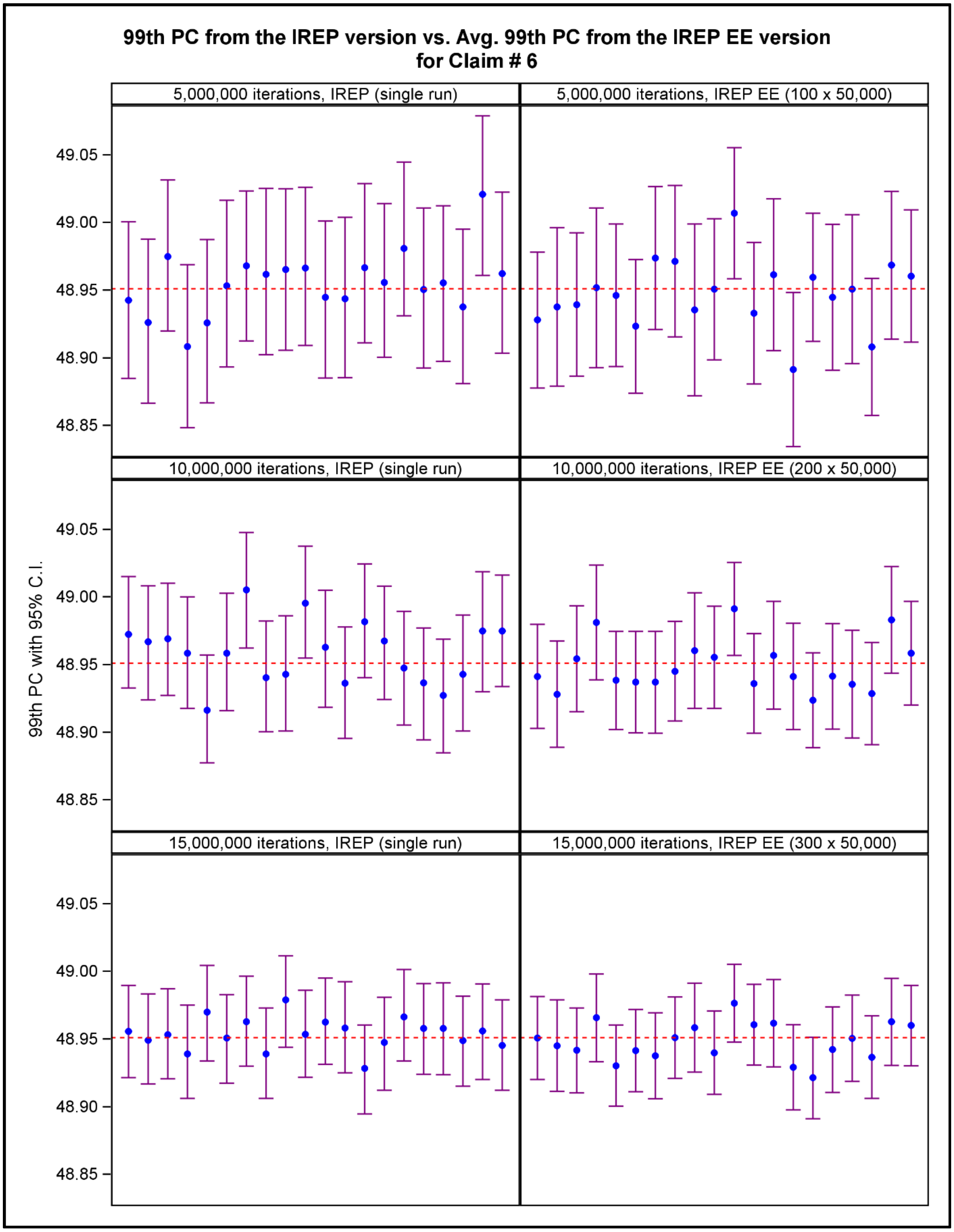


Figure 24. The 99th PC values obtained from the IREP version, versus Avg. 99th PC values from the IREP EE version, obtained at three larger number of iterations, for Claim # 6.

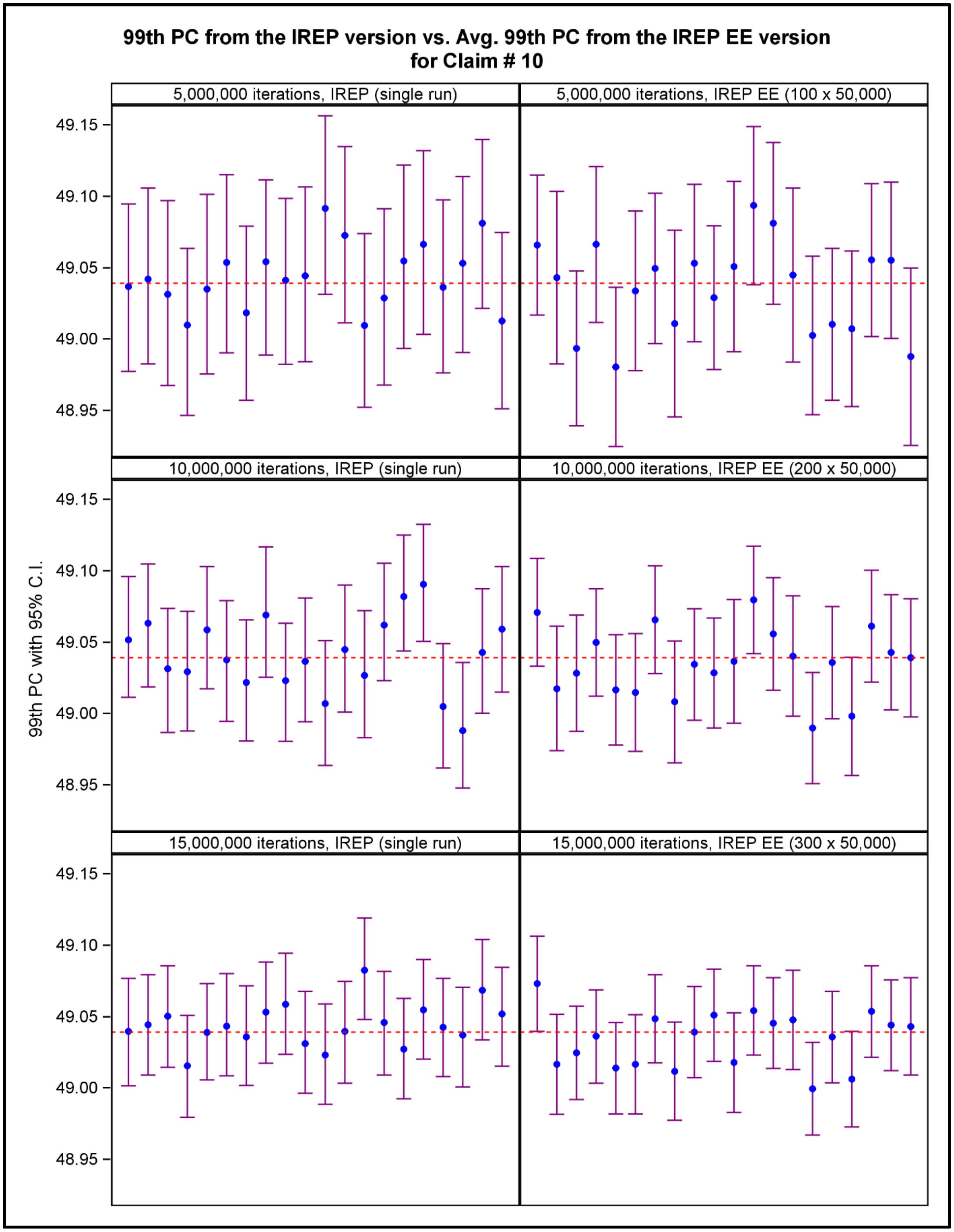


Figure 25. The 99th PC values obtained from the IREP version, versus Avg. 99th PC values from the IREP EE version, obtained at three larger number iterations, for Claim # 10.

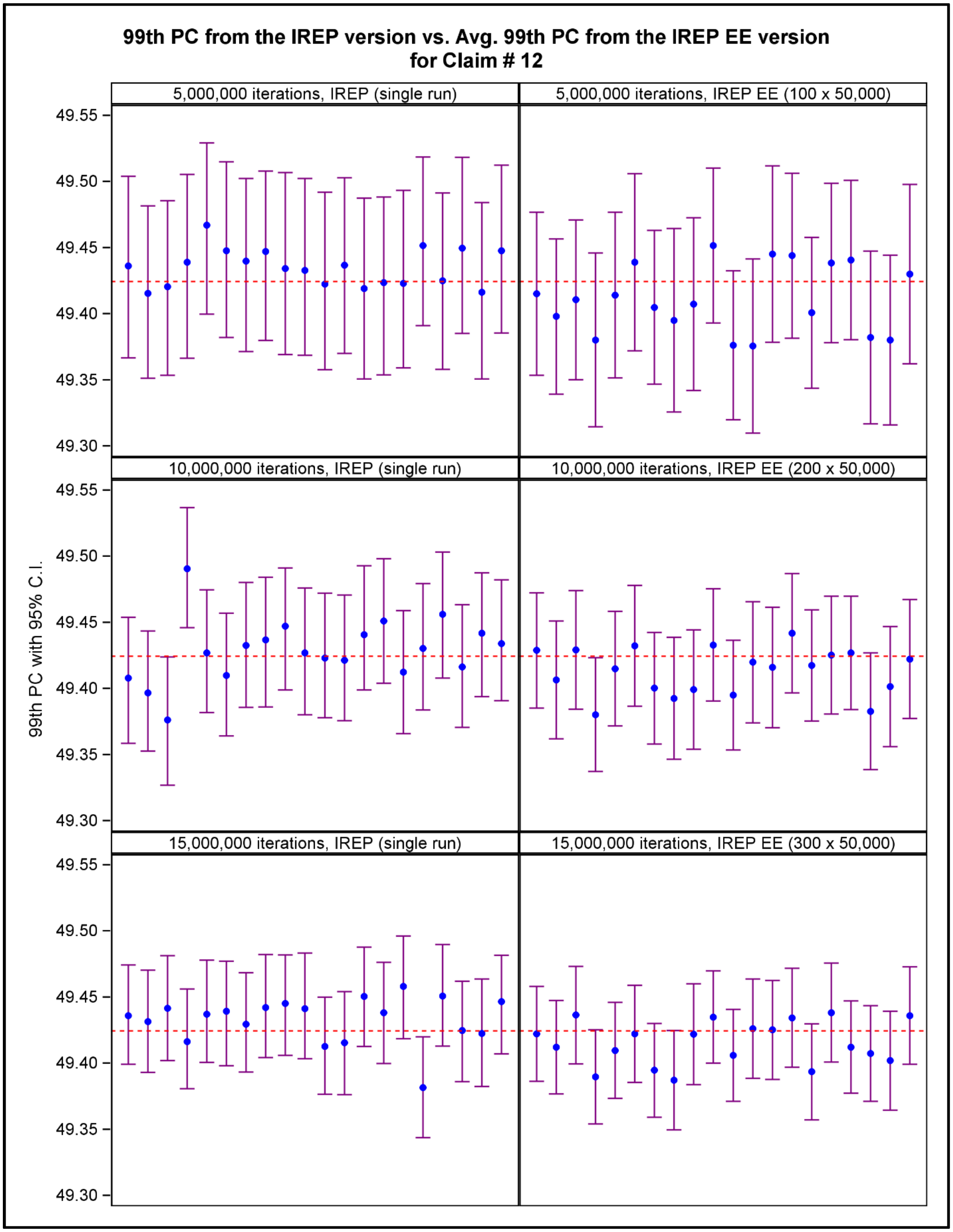


Figure 26. The 99th PC values obtained from the IREP version, versus Avg. 99th PC values from the IREP EE version, obtained at three larger number iterations, for Claim # 12.

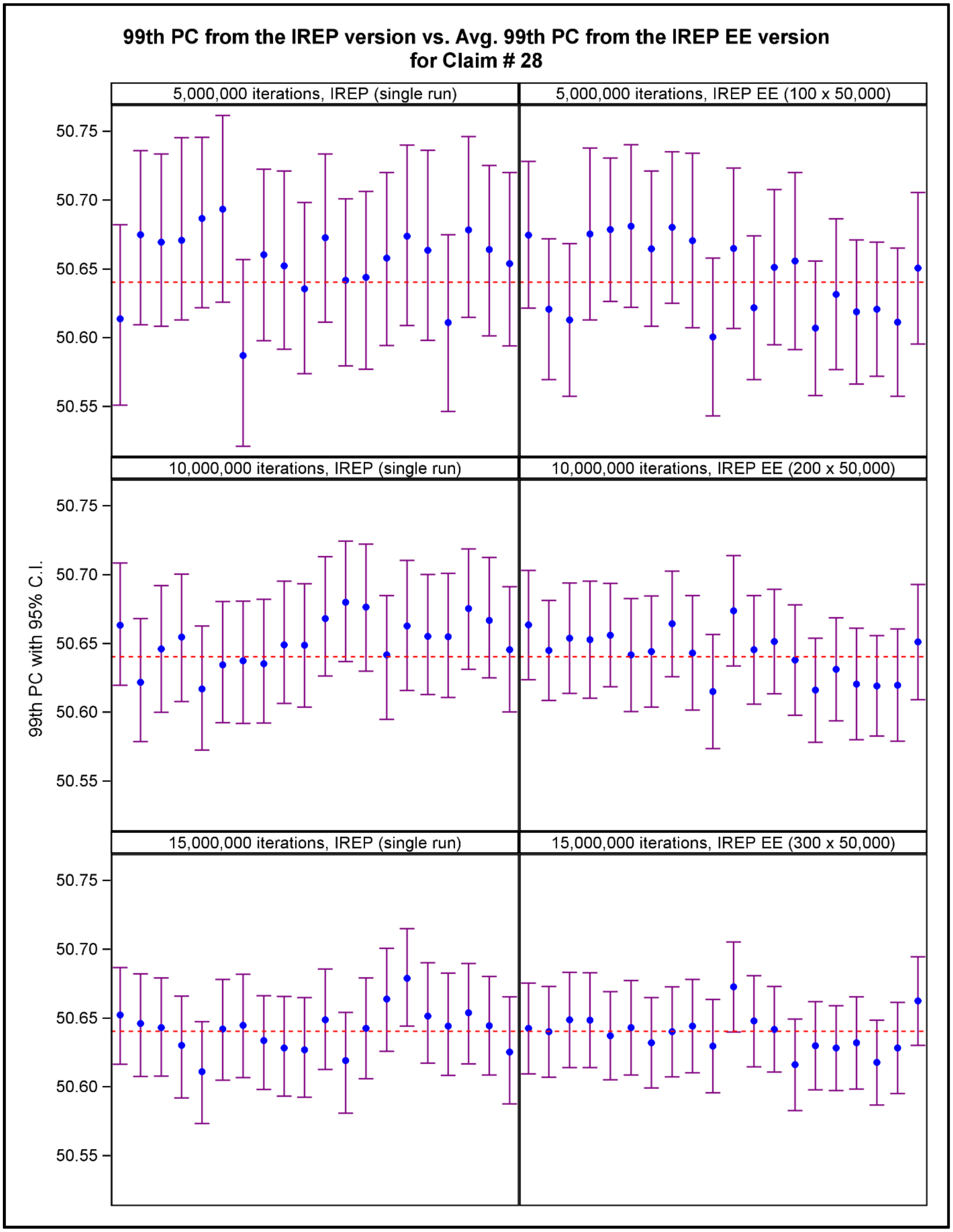


Figure 27. The 99th PC values obtained from the IREP version, versus Avg. 99th PC values from the IREP EE version, obtained at three larger number iterations, for Claim # 28.

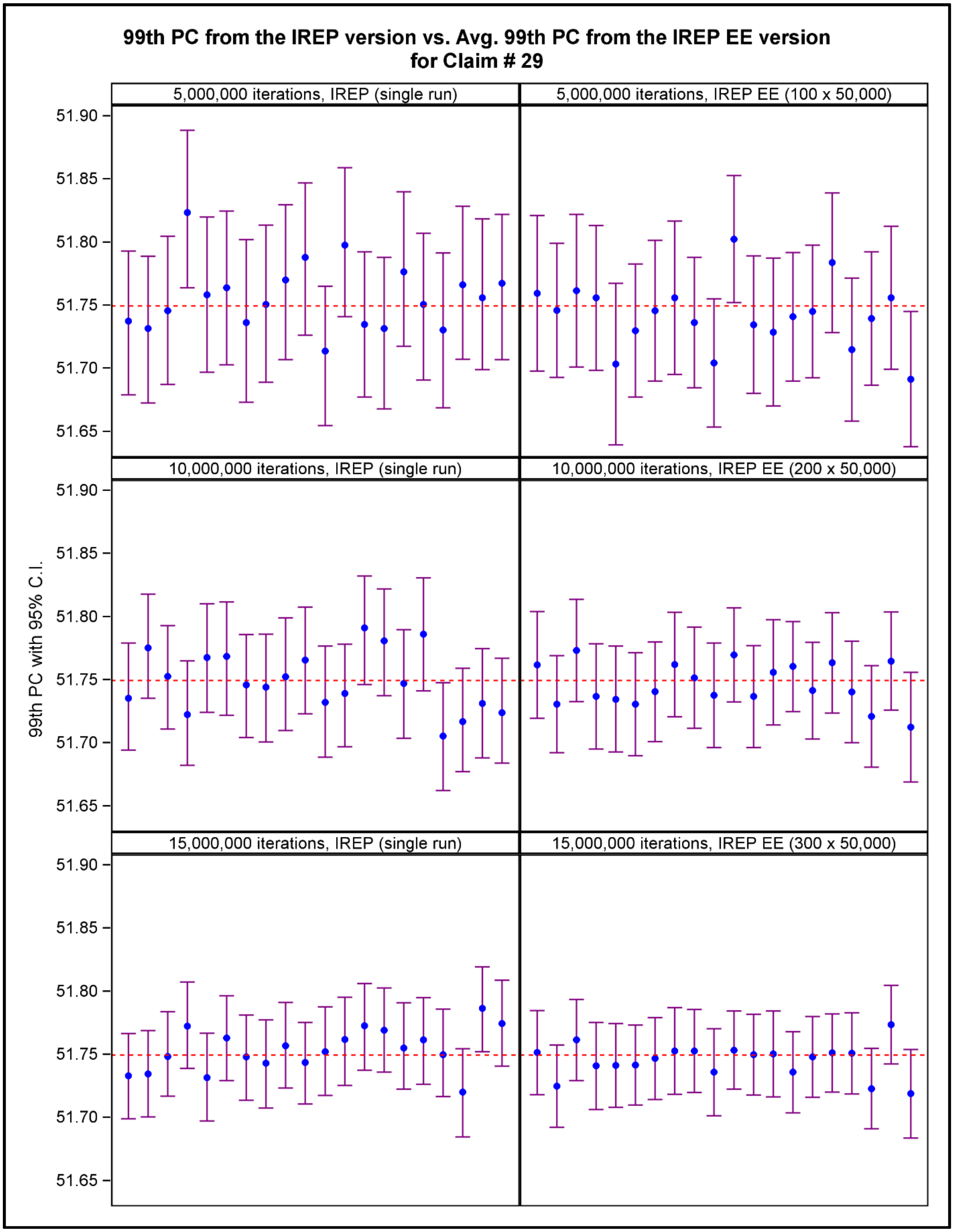


Figure 28. The 99th PC values obtained from the IREP version, versus Avg. 99th PC values from the IREP EE version, obtained at three larger number iterations, for Claim # 29.

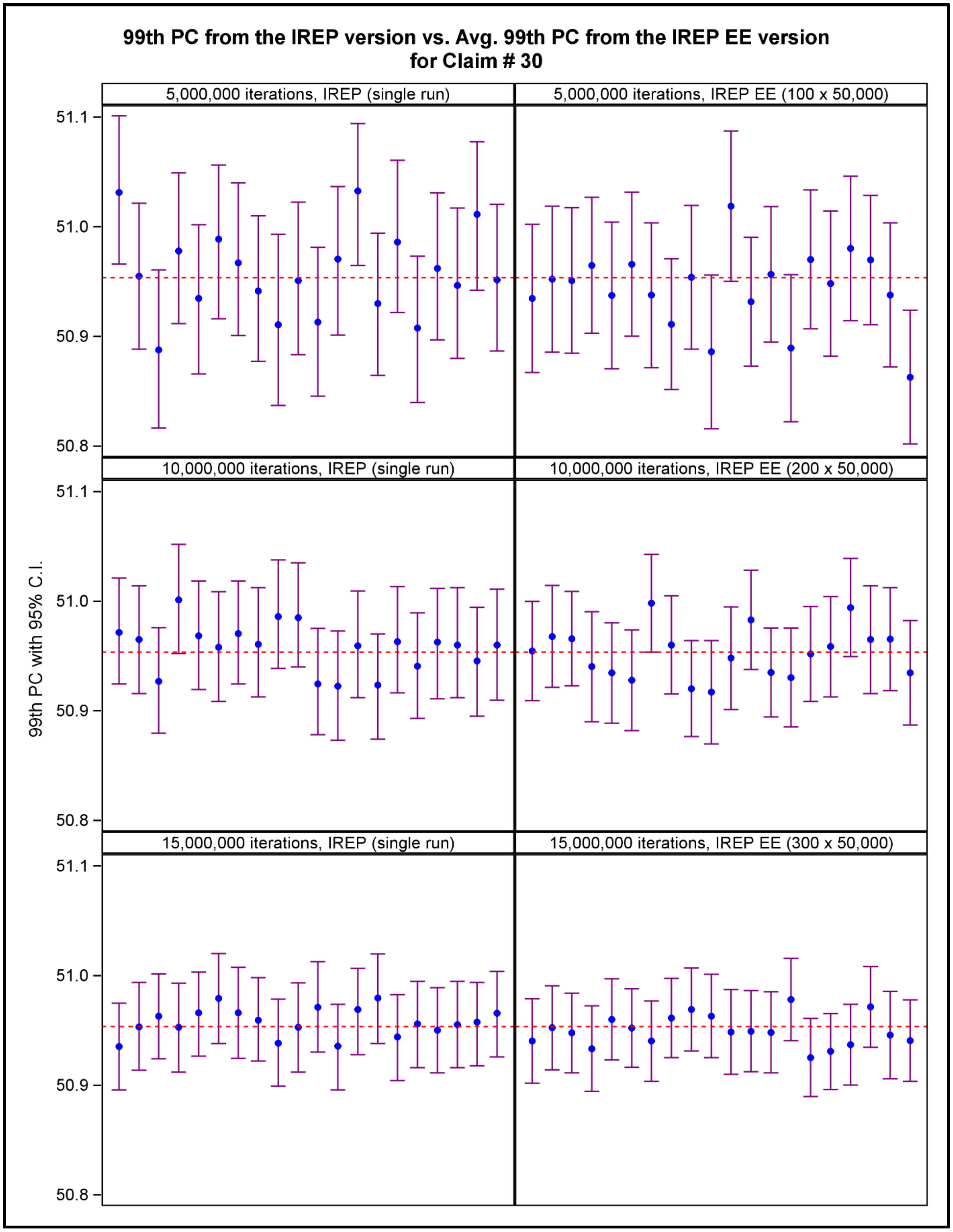


Figure 29. The 99th PC values obtained from the IREP version, versus Avg. 99th PC values from the IREP EE version, obtained at three larger number iterations, for Claim # 30.

The plots in Figures 24 through 29 also show the 95% confidence intervals around each of the 99th PC values from IREP, and around the Avg. 99th PC values from IREP EE. The lengths of the 95% confidence intervals are very similar in the two IREP versions, even though they are computed using different methods. In IREP, the 95% confidence interval for the 99th PC value is computed based on a non-parametric method, using distribution-free confidence limits for the percentiles (Meeker 2017), since the distribution of the PC values is usually not normally distributed. In IREP EE, the 95% confidence interval for the Avg. 99th PC value is computed using the normal approximation of the distribution of the 99th PC values, since the distribution of the 99th PC values from repeated runs is approximately normally distributed (Stancescu et al. 2021).

The six cases tested have 2, 4, 31, 27, 2, and 9 exposures, and because they have low number of exposures, it was possible to run them on the IREP version with such large number of iterations (5,000,000 iterations or more). However, for cases with more than 50 exposures, it is not possible to use such large number of iterations, since IREP runs out of memory for very large sample sizes. For this reason, the IREP EE version has the great advantage of being able to complete cases with any number of exposures for very large number of total iterations, by splitting a large sample size in several smaller sample sizes and averaging the results of the smaller samples. The IREP EE version can work with 30, 100, 200, 300 samples, and each of these samples can have 50,000 iterations or even more; thus, the IREP EE version can use a much larger number of total iterations than the IREP version for any case, no matter the number of exposures.

Table 1 shows the summary statistics for the 99th PC values from IREP and for the Avg. 99th PC values from IREP EE for the six cases. The summary statistics (N, Mean, Min, and Max) are shown for the previous four scenarios of total number of iterations tested: 300,000 iterations, 5,000,000 iterations, 10,000,000 iterations, and 15,000,000 iterations. For the 300,000 iterations, there were N=50 random seeds used in IREP and N=50 sets of 30 random seeds used in IREP EE for each case, and for the remaining three scenarios, there were 20 random seeds used in IREP and 20 sets of either 100, 200, or 300 random seeds used in IREP EE for each case.

Table 1. Summary statistics for the 99th PC values from IREP and for the Avg. 99th PC values from IREP EE.

Claim	Scenario	N[a]	Mean	Min	Max
6	300,000 iterations, IREP (single run)	50	48.96	48.70	49.20
6	300,000 iterations, IREP EE (30 x 10,000)	50	48.94	48.65	49.18
6	5,000,000 iterations, IREP (single run)	20	48.96	48.91	49.02
6	5,000,000 iterations, IREP EE (100 x 50,000)	20	48.95	48.89	49.01
6	10,000,000 iterations, IREP (single run)	20	48.96	48.92	49.01
6	10,000,000 iterations, IREP EE (200 x 50,000)	20	48.95	48.92	48.99
6	15,000,000 iterations, IREP (single run)	20	48.95	48.93	48.98
6	15,000,000 iterations, IREP EE (300 x 50,000)	20	48.95	48.92	48.98
10	300,000 iterations, IREP (single run)	50	49.04	48.67	49.31
10	300,000 iterations, IREP EE (30 x 10,000)	50	48.99	48.73	49.34
10	5,000,000 iterations, IREP (single run)	20	49.04	49.01	49.09
10	5,000,000 iterations, IREP EE (100 x 50,000)	20	49.04	48.98	49.09
10	10,000,000 iterations, IREP (single run)	20	49.04	48.99	49.09
10	10,000,000 iterations, IREP EE (200 x 50,000)	20	49.04	48.99	49.08
10	15,000,000 iterations, IREP (single run)	20	49.04	49.02	49.08
10	15,000,000 iterations, IREP EE (300 x 50,000)	20	49.03	49.00	49.07
12	300,000 iterations, IREP (single run)	50	49.45	49.13	49.73
12	300,000 iterations, IREP EE (30 x 10,000)	50	49.40	49.18	49.60
12	5,000,000 iterations, IREP (single run)	20	49.43	49.42	49.47
12	5,000,000 iterations, IREP EE (100 x 50,000)	20	49.41	49.38	49.45
12	10,000,000 iterations, IREP (single run)	20	49.43	49.38	49.49
12	10,000,000 iterations, IREP EE (200 x 50,000)	20	49.41	49.38	49.44
12	15,000,000 iterations, IREP (single run)	20	49.43	49.38	49.46
12	15,000,000 iterations, IREP EE (300 x 50,000)	20	49.42	49.39	49.44
28	300,000 iterations, IREP (single run)	50	50.65	50.42	50.90
28	300,000 iterations, IREP EE (30 x 10,000)	50	50.61	50.40	50.89
28	5,000,000 iterations, IREP (single run)	20	50.66	50.59	50.69
28	5,000,000 iterations, IREP EE (100 x 50,000)	20	50.64	50.60	50.68
28	10,000,000 iterations, IREP (single run)	20	50.65	50.62	50.68
28	10,000,000 iterations, IREP EE (200 x 50,000)	20	50.64	50.62	50.67
28	15,000,000 iterations, IREP (single run)	20	50.64	50.61	50.68
28	15,000,000 iterations, IREP EE (300 x 50,000)	20	50.64	50.62	50.67
29	300,000 iterations, IREP (single run)	50	51.73	51.44	52.02
29	300,000 iterations, IREP EE (30 x 10,000)	50	51.71	51.41	52.06
29	5,000,000 iterations, IREP (single run)	20	51.76	51.71	51.82

29	5,000,000 iterations, IREP EE (100 x 50,000)	20	51.74	51.69	51.80
29	10,000,000 iterations, IREP (single run)	20	51.75	51.71	51.79
29	10,000,000 iterations, IREP EE (200 x 50,000)	20	51.75	51.71	51.77
29	15,000,000 iterations, IREP (single run)	20	51.75	51.72	51.79
29	15,000,000 iterations, IREP EE (300 x 50,000)	20	51.75	51.72	51.77
30	300,000 iterations, IREP (single run)	50	50.97	50.73	51.16
30	300,000 iterations, IREP EE (30 x 10,000)	50	50.88	50.48	51.21
30	5,000,000 iterations, IREP (single run)	20	50.96	50.89	51.03
30	5,000,000 iterations, IREP EE (100 x 50,000)	20	50.94	50.86	51.02
30	10,000,000 iterations, IREP (single run)	20	50.96	50.92	51.00
30	10,000,000 iterations, IREP EE (200 x 50,000)	20	50.95	50.92	51.00
30	15,000,000 iterations, IREP (single run)	20	50.96	50.94	50.98
30	15,000,000 iterations, IREP EE (300 x 50,000)	20	50.95	50.93	50.98

[a] N = number of repetitions

On average, based on the 50 repetitions used for the 300,000 iterations scenario (corresponding to the IREP EE settings used in the IREP v.5.9), the Avg. 99th PC values from IREP EE are slightly smaller than the 99th PC values from IREP (with differences that range between 0.02 and 0.09, for the six cases). For the 5,000,000 iterations, 10,000,000 iterations, and 15,000,000 iterations scenarios, the Avg. 99th PC values from IREP EE are either identical (in 6 out of 18 scenarios) or slightly lower than the 99th PC values from IREP, based on the 20 repetitions used (the differences are either 0.01 in 7 out of 18 scenarios, or 0.02 in 5 out of 18 scenarios).

This shows that as the number of total iterations increases, there is no significant difference between the 99th PC values obtained from IREP versus the Avg. 99th PC values obtained from IREP EE, for the same number of total iterations used.

Table 2 shows the summary statistics for the lengths of the 95% confidence intervals around the 99th PC values from IREP and around the Avg. 99th PC values from IREP EE for the six cases. The summary statistics are shown for the previously used four scenarios of total numbers of iterations tested. On average, based on the 50 repetitions used for the 300,000 iterations scenario, the lengths of the 95% CIs for the Avg. 99th PC values from IREP EE are smaller than the 99th PC values from IREP (differences range between 0.03 and 0.08, for the six cases). For the 5,000,000 iterations, 10,000,000 iterations, and 15,000,000 iterations scenarios, the lengths of the 95% CIs for Avg. 99th PC values from IREP EE are slightly smaller than the lengths of the 95% CIs for the 99th PC values from IREP (in all scenarios the differences are less than 0.01), based on the 20 repetitions used.

Table 2. Summary statistics for the lengths of the 95% confidence intervals of the 99th PC values from IREP and the Avg. 99th PC values from IREP EE.

Claim	Scenario	N[a]	Mean	Min	Max
6	300,000 iterations, IREP (single run)	50	0.482	0.418	0.554
6	300,000 iterations, IREP EE (30 x 10,000)	50	0.452	0.241	0.628
6	5,000,000 iterations, IREP (single run)	20	0.117	0.111	0.123
6	5,000,000 iterations, IREP EE (100 x 50,000)	20	0.107	0.095	0.127
6	10,000,000 iterations, IREP (single run)	20	0.084	0.080	0.089
6	10,000,000 iterations, IREP EE (200 x 50,000)	20	0.077	0.069	0.086
6	15,000,000 iterations, IREP (single run)	20	0.067	0.064	0.071
6	15,000,000 iterations, IREP EE (300 x 50,000)	20	0.062	0.057	0.067
10	300,000 iterations, IREP (single run)	50	0.493	0.404	0.601
10	300,000 iterations, IREP EE (30 x 10,000)	50	0.457	0.299	0.610
10	5,000,000 iterations, IREP (single run)	20	0.123	0.116	0.130
10	5,000,000 iterations, IREP EE (100 x 50,000)	20	0.112	0.098	0.131
10	10,000,000 iterations, IREP (single run)	20	0.086	0.081	0.091
10	10,000,000 iterations, IREP EE (200 x 50,000)	20	0.080	0.075	0.087
10	15,000,000 iterations, IREP (single run)	20	0.071	0.067	0.075
10	15,000,000 iterations, IREP EE (300 x 50,000)	20	0.066	0.062	0.070
12	300,000 iterations, IREP (single run)	50	0.549	0.464	0.658
12	300,000 iterations, IREP EE (30 x 10,000)	50	0.514	0.346	0.675
12	5,000,000 iterations, IREP (single run)	20	0.133	0.127	0.139
12	5,000,000 iterations, IREP EE (100 x 50,000)	20	0.125	0.113	0.139
12	10,000,000 iterations, IREP (single run)	20	0.094	0.091	0.098
12	10,000,000 iterations, IREP EE (200 x 50,000)	20	0.088	0.083	0.092
12	15,000,000 iterations, IREP (single run)	20	0.077	0.073	0.081
12	15,000,000 iterations, IREP EE (300 x 50,000)	20	0.073	0.070	0.076
28	300,000 iterations, IREP (single run)	50	0.527	0.459	0.618
28	300,000 iterations, IREP EE (30 x 10,000)	50	0.448	0.317	0.590
28	5,000,000 iterations, IREP (single run)	20	0.128	0.122	0.138
28	5,000,000 iterations, IREP EE (100 x 50,000)	20	0.111	0.097	0.129
28	10,000,000 iterations, IREP (single run)	20	0.090	0.087	0.095
28	10,000,000 iterations, IREP EE (200 x 50,000)	20	0.079	0.073	0.085
28	15,000,000 iterations, IREP (single run)	20	0.073	0.068	0.078
28	15,000,000 iterations, IREP EE (300 x 50,000)	20	0.066	0.062	0.069
29	300,000 iterations, IREP (single run)	50	0.493	0.406	0.555
29	300,000 iterations, IREP EE (30 x 10,000)	50	0.449	0.315	0.537
29	5,000,000 iterations, IREP (single run)	20	0.120	0.110	0.129

29	5,000,000 iterations, IREP EE (100 x 50,000)	20	0.111	0.100	0.128
29	10,000,000 iterations, IREP (single run)	20	0.085	0.081	0.090
29	10,000,000 iterations, IREP EE (200 x 50,000)	20	0.080	0.071	0.087
29	15,000,000 iterations, IREP (single run)	20	0.068	0.065	0.070
29	15,000,000 iterations, IREP EE (300 x 50,000)	20	0.065	0.062	0.070
30	300,000 iterations, IREP (single run)	50	0.570	0.507	0.655
30	300,000 iterations, IREP EE (30 x 10,000)	50	0.520	0.362	0.653
30	5,000,000 iterations, IREP (single run)	20	0.137	0.130	0.156
30	5,000,000 iterations, IREP EE (100 x 50,000)	20	0.129	0.118	0.140
30	10,000,000 iterations, IREP (single run)	20	0.098	0.094	0.101
30	10,000,000 iterations, IREP EE (200 x 50,000)	20	0.091	0.081	0.100
30	15,000,000 iterations, IREP (single run)	20	0.079	0.076	0.083
30	15,000,000 iterations, IREP EE (300 x 50,000)	20	0.074	0.069	0.080

[a] Number of repetitions

The statistics in Table 2 show another advantage of using the IREP EE version, which produce slightly better precision (smaller lengths for the 95% CIs) in the Avg. 99th PC values obtained from IREP EE, versus the 99th PC values from IREP, for the same number of total iterations used.

7.0 CONCLUSIONS

The convergence of the 99th PC values as well as of the Avg. 99th PC values, as the number of iterations increases was shown for several cases with PC values between 49.00% and 50.10%, and for each of these cases there is a clear trend for both quantities to converge to a stable value that represents the theoretical 99th percentile of the distribution of PC values for that case.

The theoretical 99th PC value of the distribution of PC values for a case cannot be determined exactly, but a 95% confidence interval for the theoretical 99th PC value can be estimated. The larger the number of iterations used in the computation, the smaller the length of the 95% confidence interval for the theoretical 99th PC value will be, and for a sample size large enough, a length for the 95% confidence interval of less than 0.15 or even less than 0.10 can be achieved.

Most of the 95% confidence intervals for the theoretical 99th PC values for the 23 tested cases do not extend above the compensability threshold of 50.0%, except for two cases for which the 95% confidence intervals for the theoretical 99th PC values includes the 50.0% threshold. The length of the 95% confidence intervals for the theoretical 99th PC value for these cases obtained from very large sample sizes is less than 0.10, and this suggests there is a limit of the precision with which the theoretical 99th PC value for a case can be bounded.

In order to speed up the convergence of the PC values and to reduce the length of the 95% confidence intervals for the Avg. 99th PC values, four different scenarios with different numbers of runs and iterations were tested, and it was concluded that the lengths of the 95% confidence intervals for the Avg. 99th PC values are smaller than 0.15 for all tested cases in Scenario 2 (200 runs with 50,000 iterations) and Scenario 3 (300 runs with 50,000 iterations), and less than 0.10 for all cases in Scenario 4 (300 runs with 100,000 iterations). The largest IREP running time observed for the 23 cases using Scenario 2 is around 2 hours, using Scenario 3 is around 3 hours, and using Scenario 4 is slightly less than 5.7 hours.

A more efficient way to reduce the length of the 95% confidence intervals for the Avg. 99th PC values and to also minimize the IREP EE running times, is to use the IREP Predictive Tool for each case to determine the minimum the number of iterations needed to achieve the desired precision for the length of the 95% confidence interval of the Avg. 99th PC values. For the 0.15 precision level, the number of iterations needed for the 23 cases with Avg. 99th PC values between 49.00% and 50.01%, varies between 5,000 and 35,000 iterations, and for the 0.10 precision level, the number of iterations needed for the 23 cases varies between 15,000 and 60,000 iterations. The largest IREP running time observed for the 23 cases when the precision level was 0.15 is less than one hour, and the largest IREP running time observed when the precision level was 0.10 was around 1.75 hours.

The precision reached for the 95% confidence intervals of the Avg. 99th PC values, from the IREP EE scenarios tested, and the largest IREP EE running times obtained for the 23 tested cases are shown in Table 3. Any of these five combinations of runs and iterations can be used in practice to achieve a desired precision level for the Avg. 99th PC values.

Table 3. Precision reached for the 95% confidence intervals of the Avg. 99th PC values, and the largest IREP EE runtimes obtained for the 23 tested cases.

IREP EE Scenario	# Runs	# Iterations	Length of 95% C.I. of Avg. 99th PC	IREP EE Runtime
Scenario 2	200	Fixed (50,000)	< 0.15	< 2.1 hours
Scenario 3	300	Fixed (50,000)	< 0.15	< 3.1 hours
Scenario 4	300	Fixed (100,000)	< 0.10	< 5.7 hours
IREP Predictive Tool	300	Variable (5,000 - 35,000)	<= 0.15	< 1 hour
IREP Predictive Tool	300	Variable (15,000 - 60,000)	<= 0.10	< 1.75 hours

For the cases with Avg. 99th PC values between 49.50% and 50.50%, one of the 5 scenarios with different numbers of runs and iterations described in Table 3 is suggested to be used to reduce the variability of the Avg. 99th PC values that is occurring by changing the sets of random seeds, and to increase the accuracy of the Avg. 99th PC value.

By increasing the number of runs in the IREP Enterprise Edition from 30 runs to 300 runs, and by increasing the number of iterations per run, the length of the 95% confidence interval of the Avg. 99th PC values can be reduced below one of the desirable precision levels of 0.15 or 0.10.

A validation exercise carried out by comparing the 99th PC values obtained from the IREP version and the Avg. 99th PC values obtained from the IREP EE version was also addressed in this paper. As the number of total iterations increases, there is no significant difference between the 99th PC values obtained from IREP versus the Avg. 99th PC values obtained from IREP EE, for the same total number of iterations used. There is slightly better precision (smaller lengths for the 95% CIs) in the Avg. 99th PC values obtained from IREP EE, versus the 99th PC values from IREP, for the same number of total iterations. The IREP EE version also has the advantage that it can work with either 30 or 300 samples, and each of these samples can have 50,000 iterations or more; thus, the IREP EE version can use a much larger number of total iterations than the IREP version for any case, regardless of the number of exposures.

REFERENCES

Meeker W. Q., Hahn G. J., Escobar L. A., Statistical Intervals: A Guide for Practitioners and Researchers, Second Edition, Wiley, 2017.

Stancescu D., Apostoaei A. I., Thomas B.A., Effect of Alternative Percentile Definition on PC values, NIOSH, October 2021.

LIST OF TERMS, ABBREVIATIONS AND ACRONYMS

Distribution types

LN(GM, GSD)	Lognormal probability distribution defined by parameters GM and GSD
GM	Geometric Mean
GSD	Geometric Standard Deviation
N(mean, SD)	Normal probability distribution defined by parameters mean and SD
mean	Mean of a Normal (Gaussian) probability distribution
SD	Standard deviation

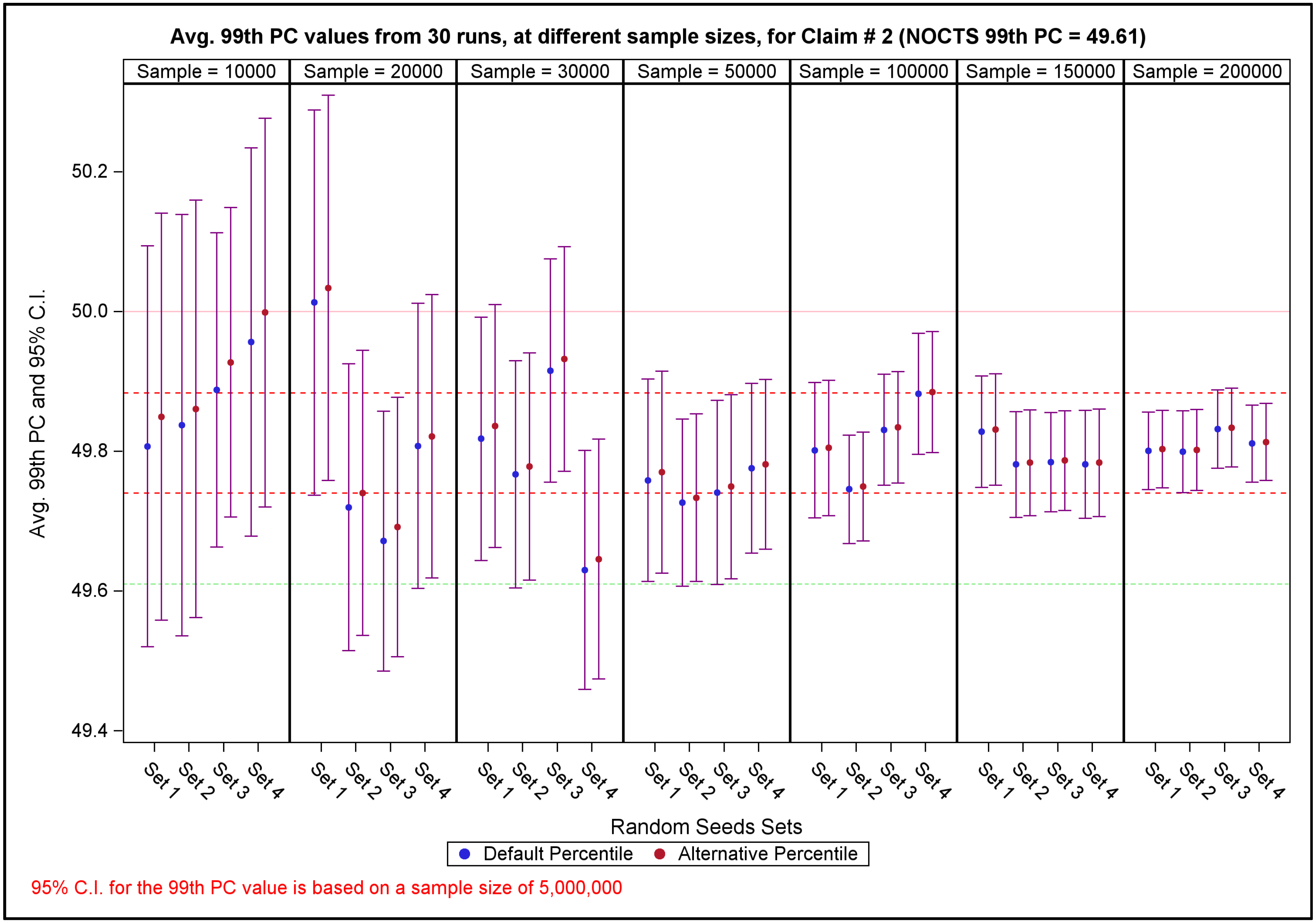
Other terms

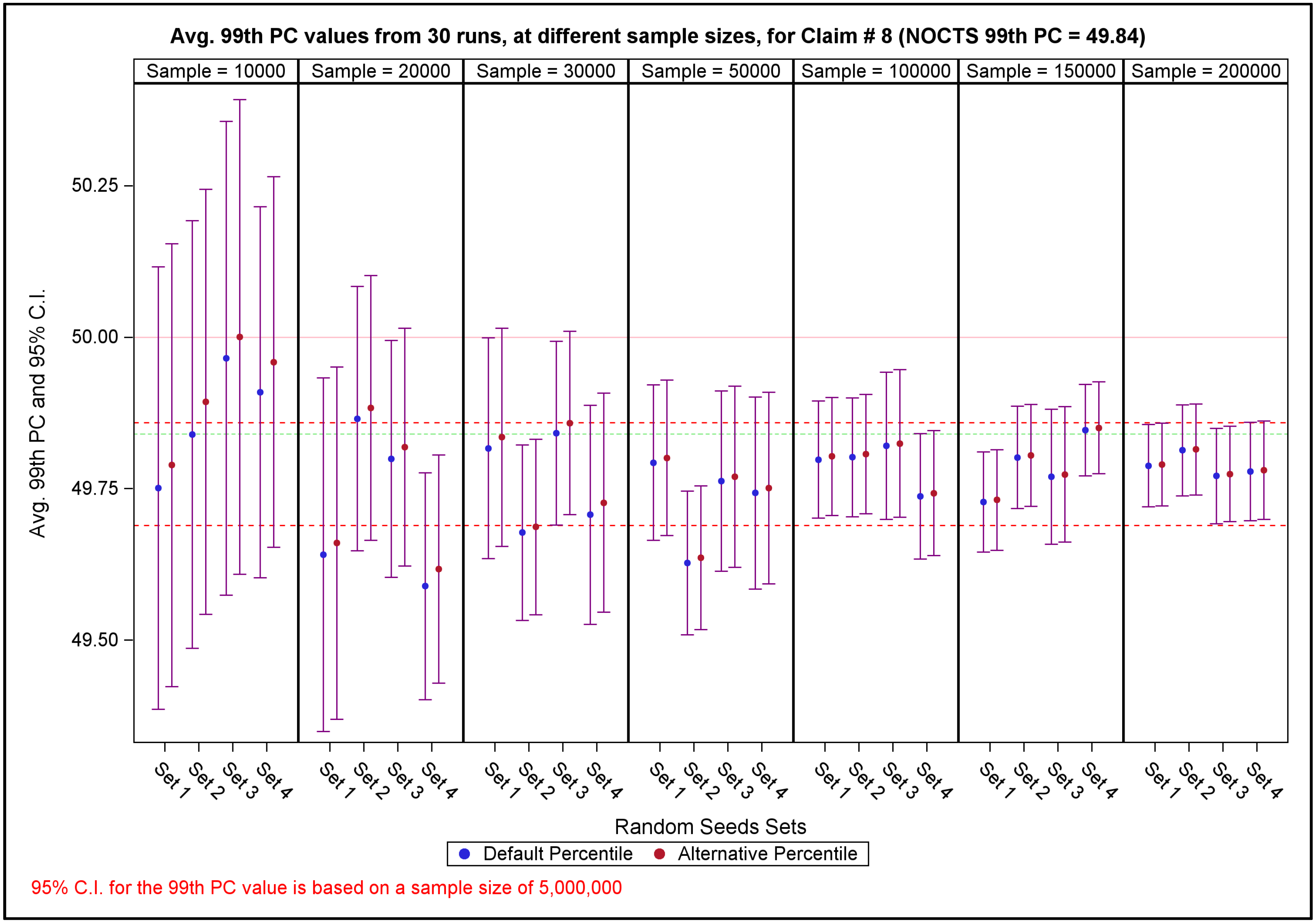
99th PC value	99th percentile of probability of causation (PC)
Avg. 99th PC value	Average of the 99th percentiles of PC obtained in sets of 30, 100, 200, or 300 runs.
ABRWH	Advisory Board on Radiation and Worker Health
IREP	Interactive Radio-Epidemiological Program
DCAS	Division of Compensation Analysis and Support
NIOSH	National Institute of Occupational Safety and Health
NOCTS	NIOSH DCAS Claims Tracking System
PC	Probability of Causation

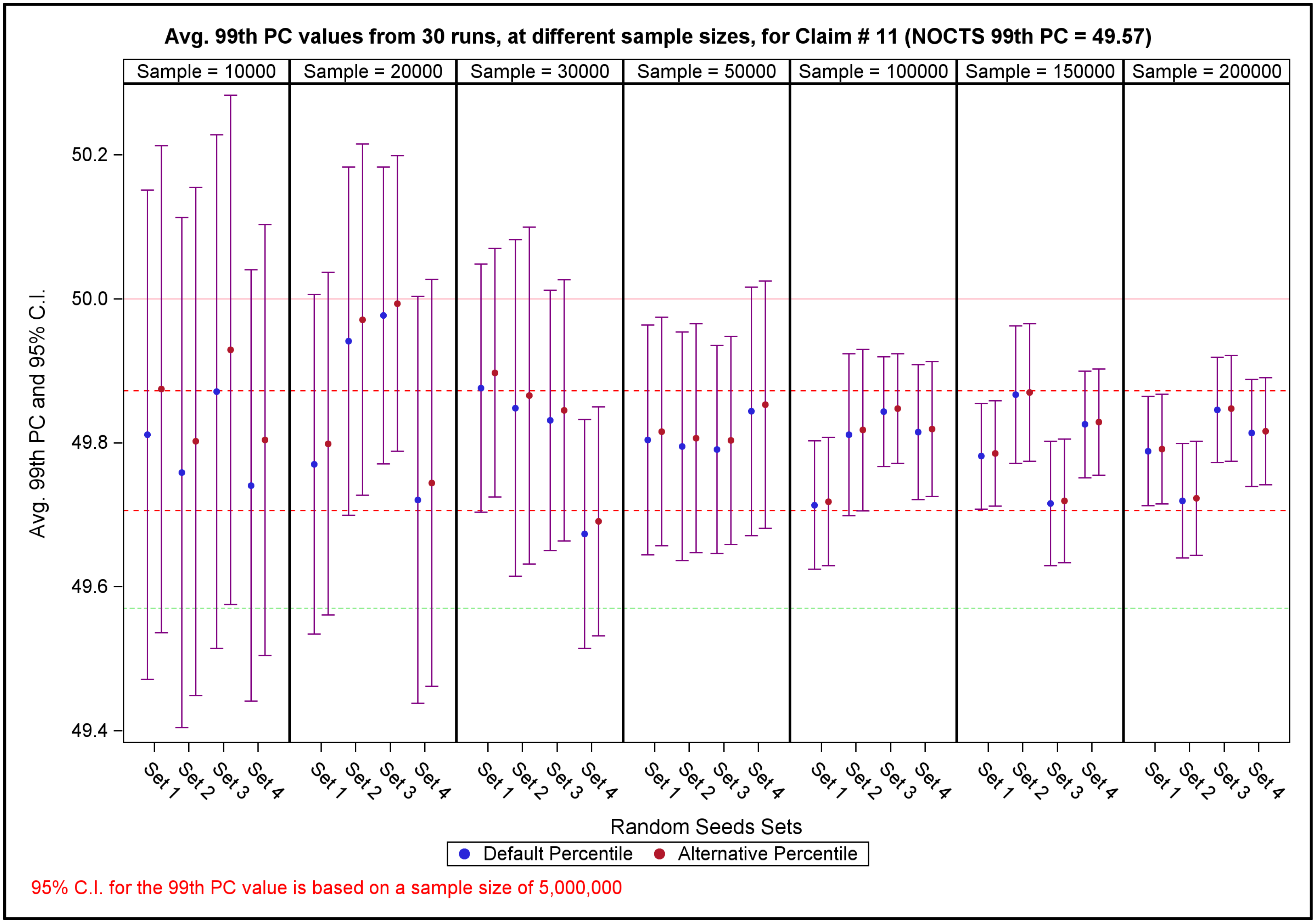
APPENDIX A

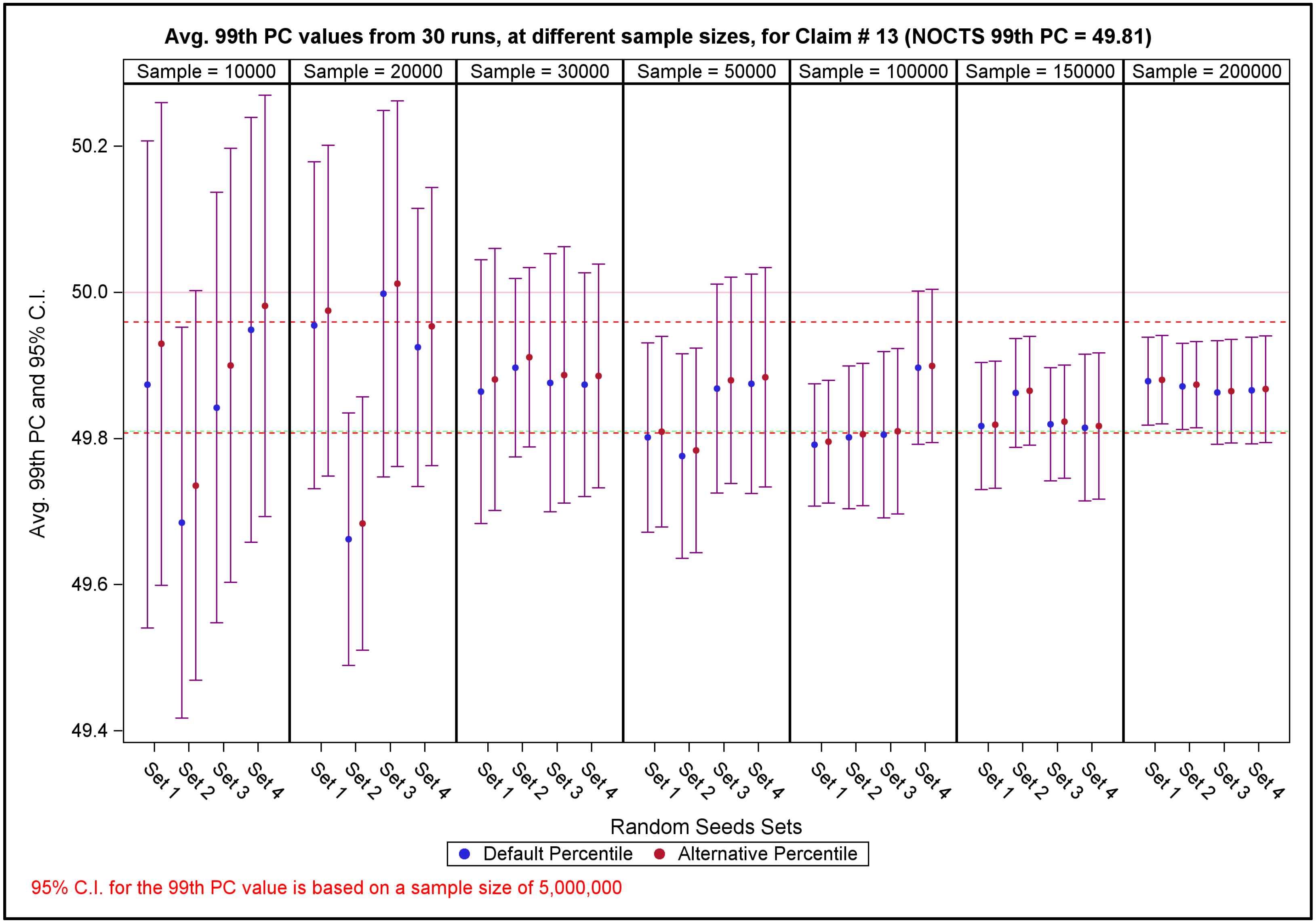
This appendix presents plots showing the convergence of the Avg. 99th PC values for 11 cases, obtained at 7 different sample sizes (10000, 20000, 30000, 50000, 100000, 150000, and 200000 iterations). For each sample size, four sets of 30 random seeds were used (the same four sets of 30 random seeds were used at each sample size).

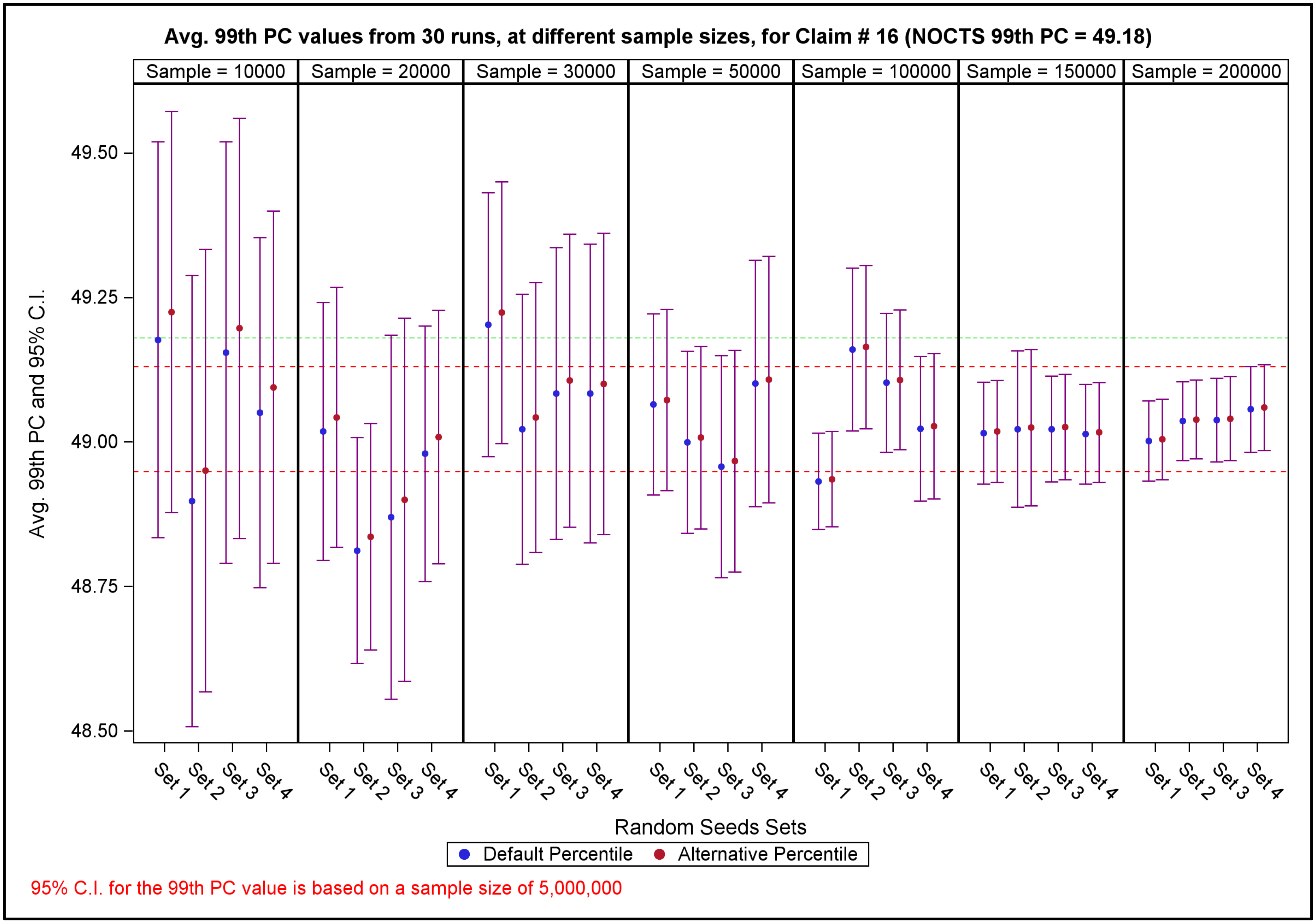
The 95% confidence interval for the theoretical 99th PC value for each case was obtained using a non-parametric method (Meeker 2017) either directly in the IREP version, or by combining multiple data sets with individual PC values obtained from IREP, and then using the non-parametric method for the combined data set with all the individual PC values.


Avg. 99th PC values from 30 runs, at different sample sizes, for Claim # 2 (NOCTS 99th PC = 49.61)
Sample = 10000
Sample = 20000
Sample = 30000
Sample = 50000
Sample = 100000
Sample = 150000
Sample = 200000
Avg. 99th PC and 95% C.I.
50.2
50.0
49.8
49.6
49.4
Set 1
Set 2
Set 3
Set 4
Random Seeds Sets
Default Percentile
Alternative Percentile
95% C.I. for the 99th PC value is based on a sample size of 5,000,000

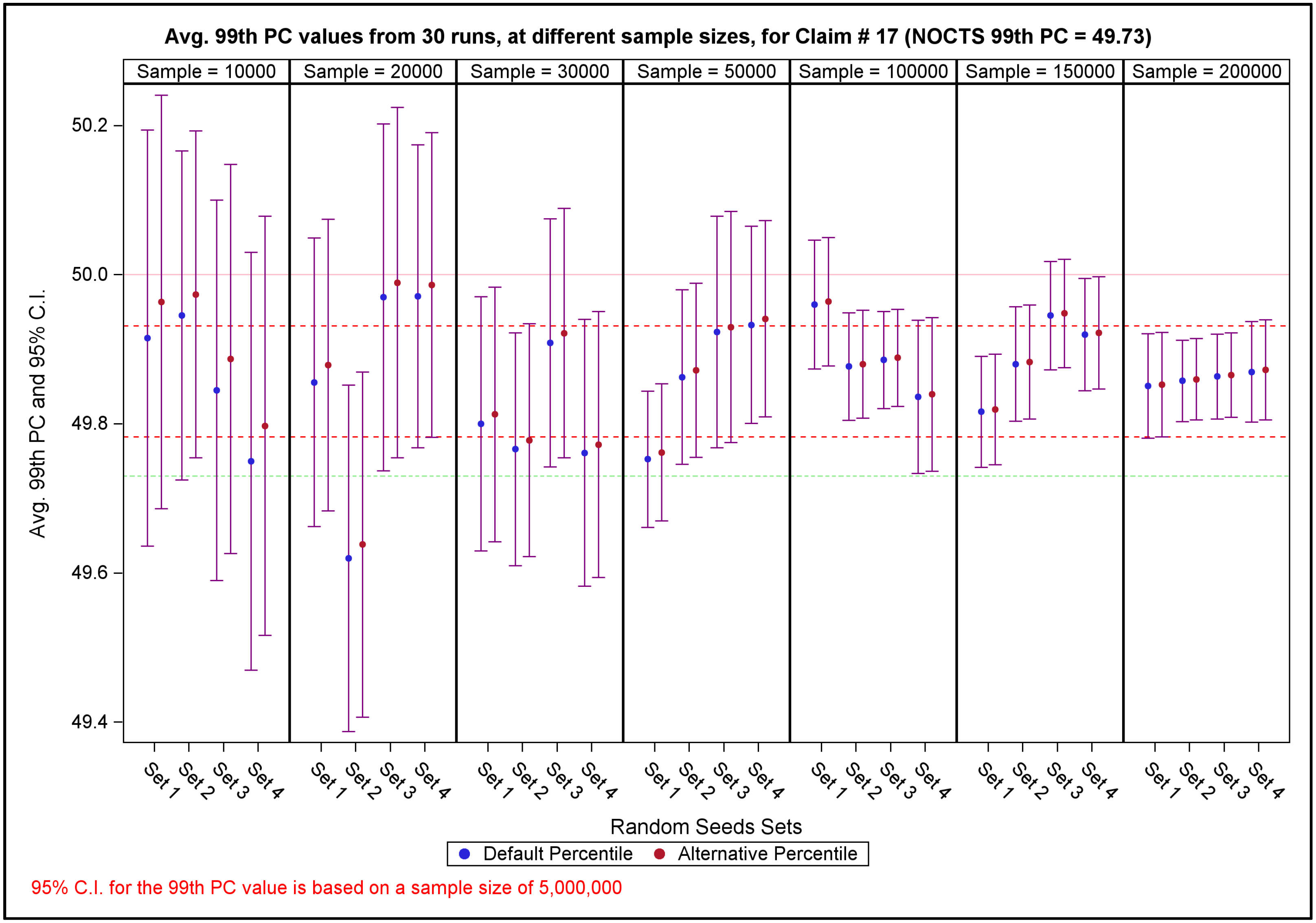

Avg. 99th PC values from 30 runs, at different sample sizes, for Claim # 8 (NOCTS 99th PC = 49.84)
Sample = 10000
Sample = 20000
Sample = 30000
Sample = 50000
Sample = 100000
Sample = 150000
Sample = 200000
Avg. 99th PC and 95% C.I.
50.25
50.00
49.75
49.50
Set 1
Set 2
Set 3
Set 4
Random Seeds Sets
Default Percentile
Alternative Percentile
95% C.I. for the 99th PC value is based on a sample size of 5,000,000


Avg. 99th PC values from 30 runs, at different sample sizes, for Claim # 11 (NOCTS 99th PC = 49.57)
Sample = 10000
Sample = 20000
Sample = 30000
Sample = 50000
Sample = 100000
Sample = 150000
Sample = 200000
Avg. 99th PC and 95% C.I.
50.2
50.0
49.8
49.6
49.4
Set 1
Set 2
Set 3
Set 4
Random Seeds Sets
Default Percentile
Alternative Percentile
95% C.I. for the 99th PC value is based on a sample size of 5,000,000

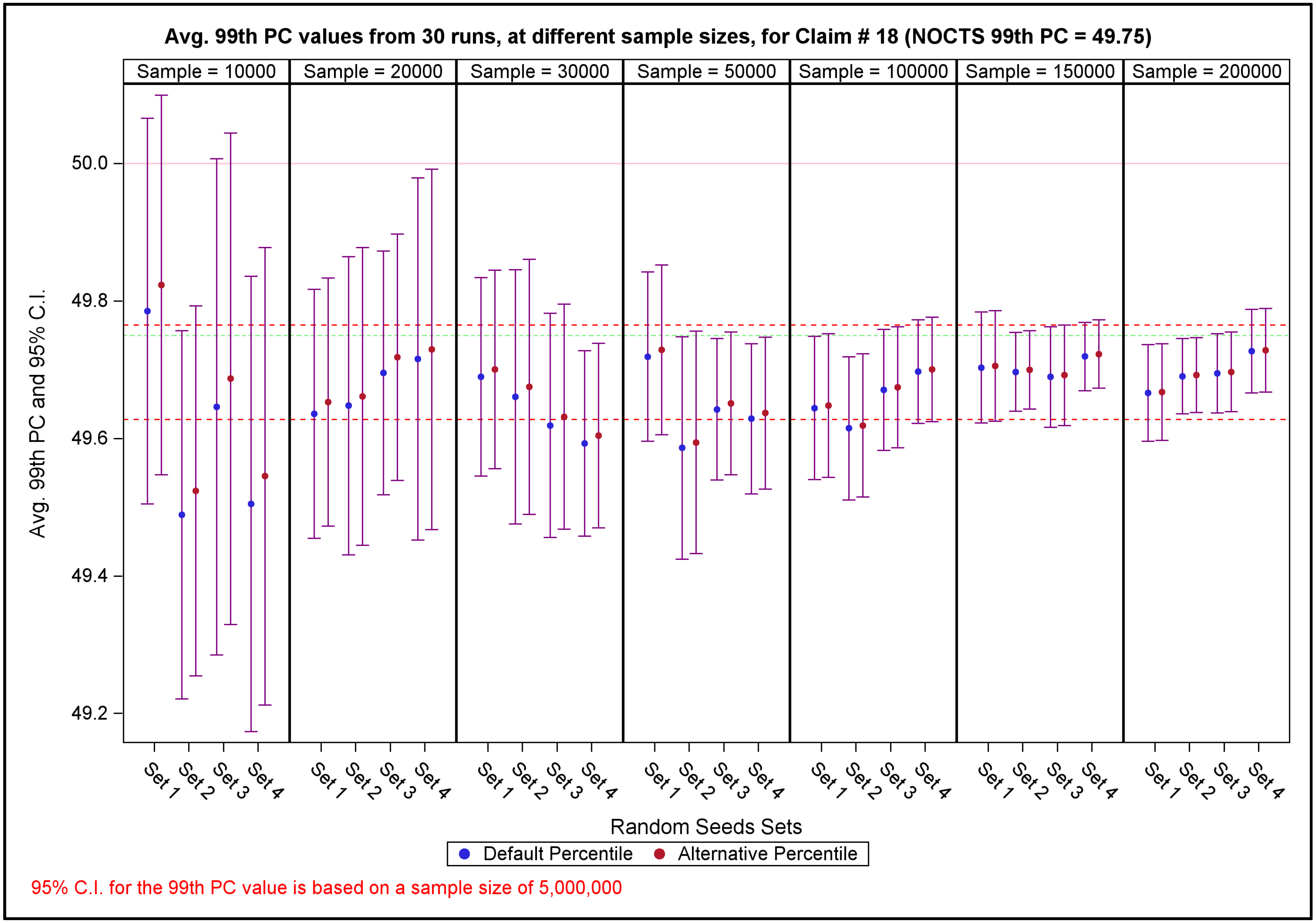

Avg. 99th PC values from 30 runs, at different sample sizes, for Claim # 13 (NOCTS 99th PC = 49.81)
Sample = 10000
Sample = 20000
Sample = 30000
Sample = 50000
Sample = 100000
Sample = 150000
Sample = 200000
Avg. 99th PC and 95% C.I.
50.2
50.0
49.8
49.6
49.4
Set 1
Set 2
Set 3
Set 4
Random Seeds Sets
Default Percentile
Alternative Percentile
95% C.I. for the 99th PC value is based on a sample size of 5,000,000

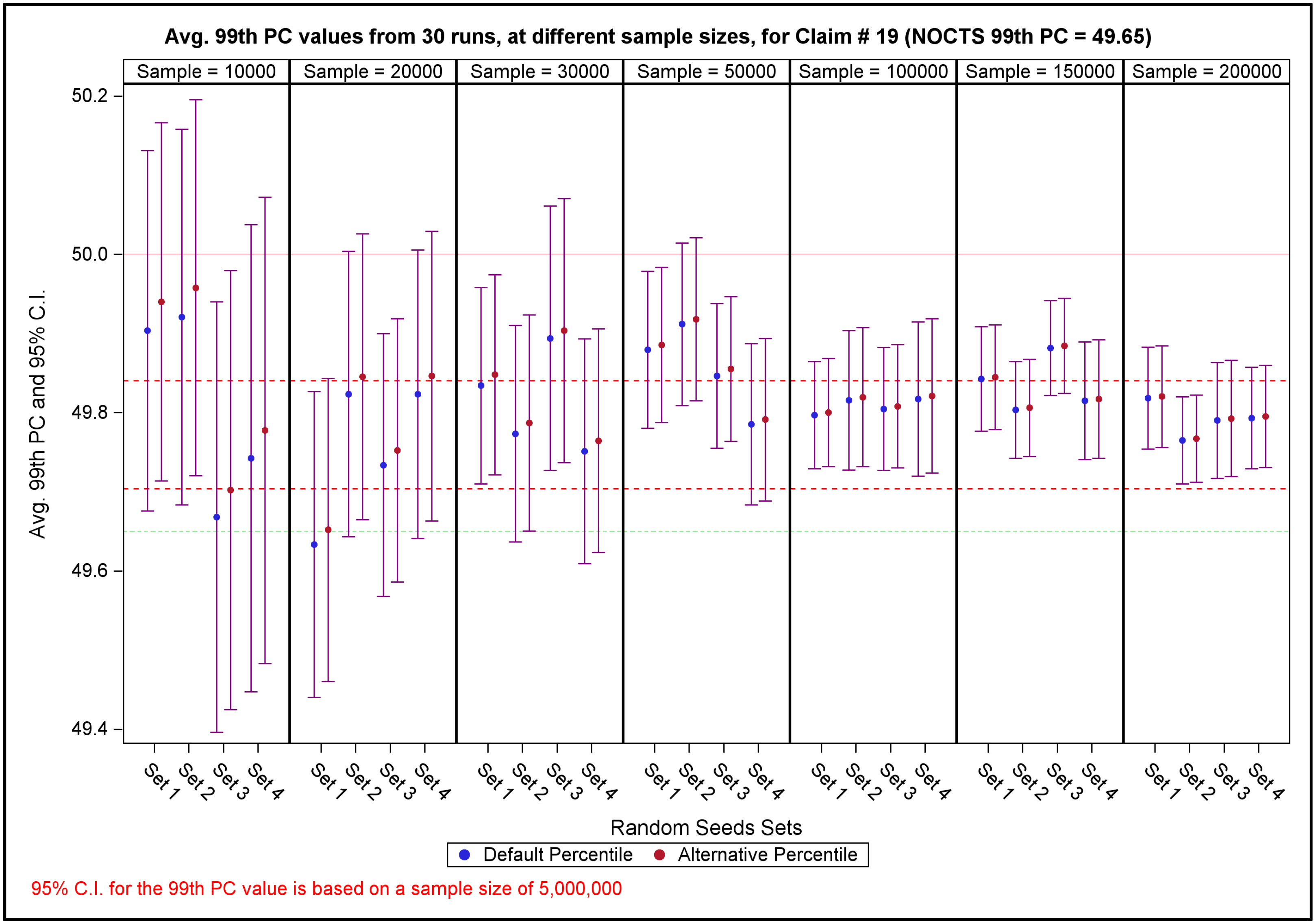


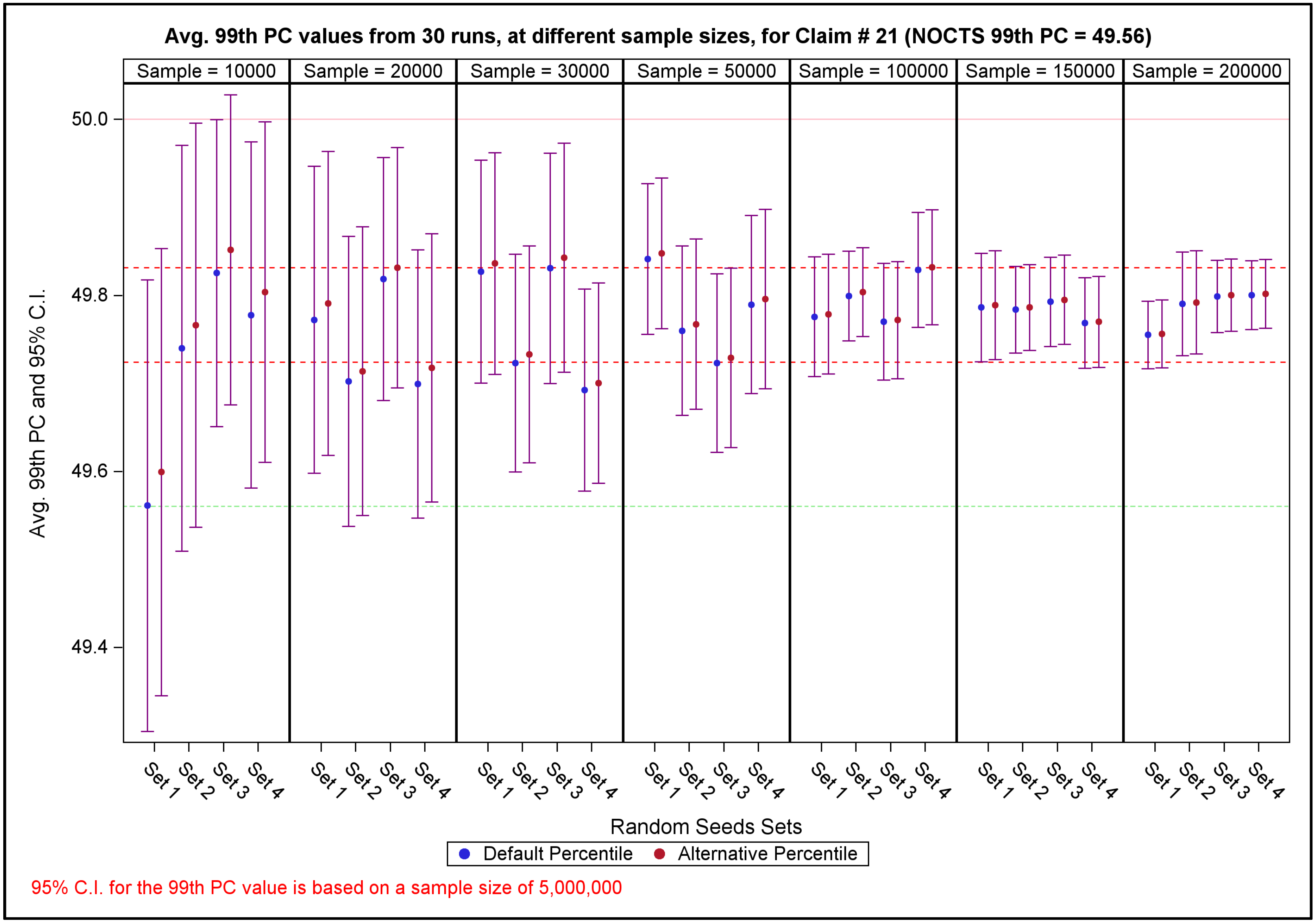
Avg. 99th PC values from 30 runs, at different sample sizes, for Claim # 16 (NOCTS 99th PC = 49.18)
Sample = 10000
Sample = 20000
Sample = 30000
Sample = 50000
Sample = 100000
Sample = 150000
Sample = 200000
Avg. 99th PC and 95% C.I.
49.50
49.25
49.00
48.75
48.50
Set 1
Set 2
Set 3
Set 4
Random Seeds Sets
Default Percentile
Alternative Percentile
95% C.I. for the 99th PC value is based on a sample size of 5,000,000

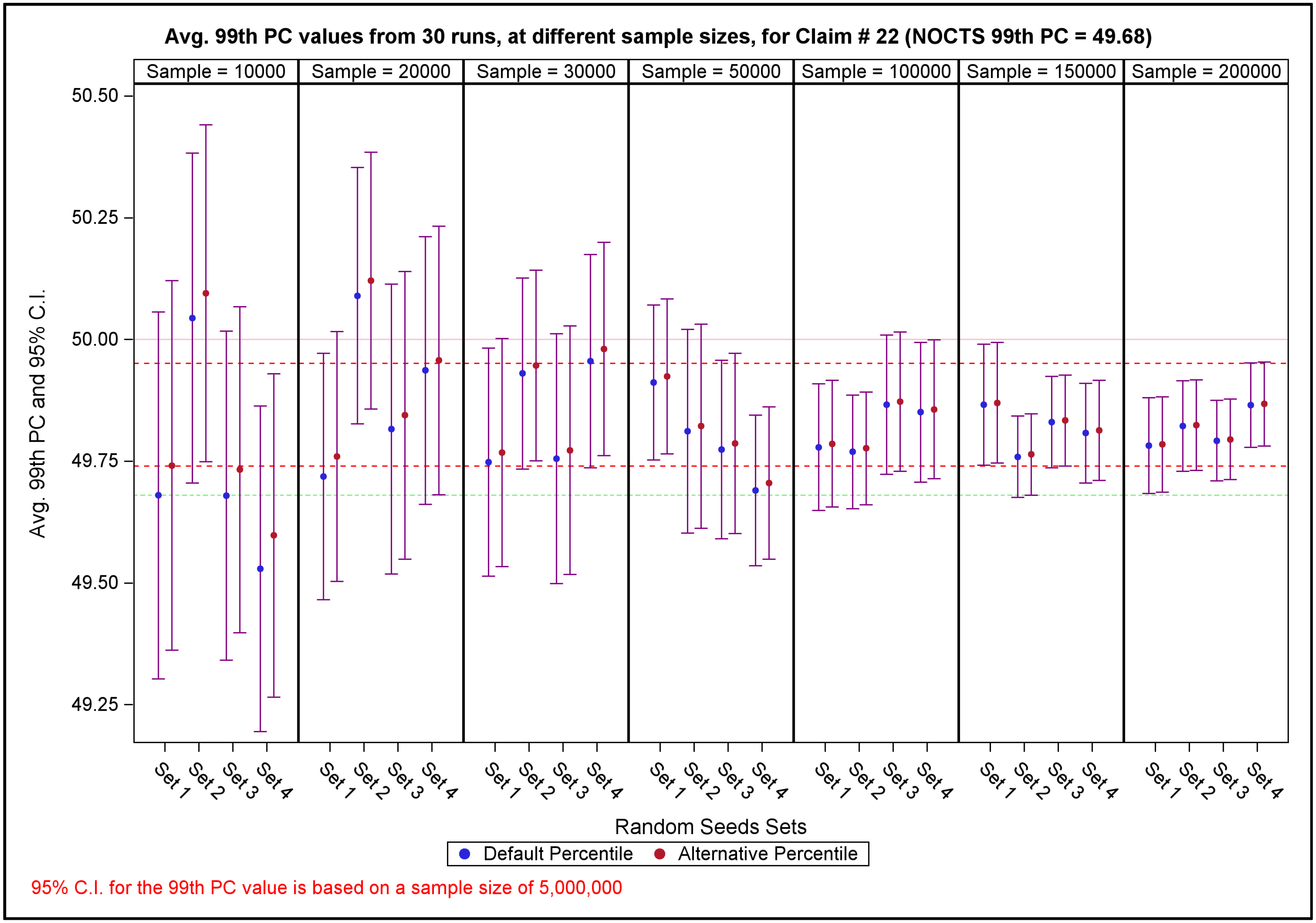


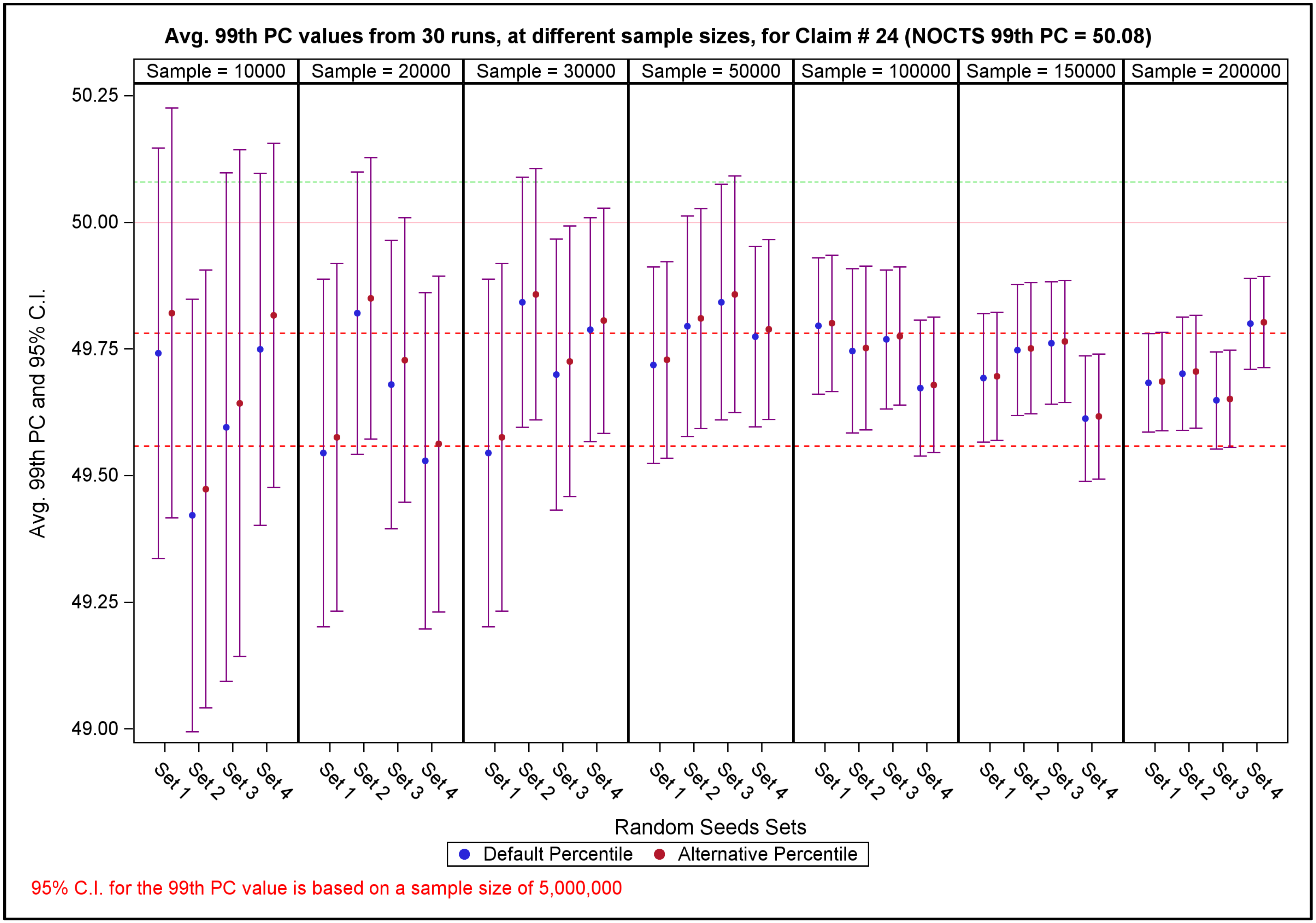
Avg. 99th PC values from 30 runs, at different sample sizes, for Claim # 17 (NOCTS 99th PC = 49.73)
Sample = 10000
Sample = 20000
Sample = 30000
Sample = 50000
Sample = 100000
Sample = 150000
Sample = 200000
Avg. 99th PC and 95% C.I.
50.2
50.0
49.8
49.6
49.4
Set 1
Set 2
Set 3
Set 4
Random Seeds Sets
Default Percentile
Alternative Percentile
95% C.I. for the 99th PC value is based on a sample size of 5,000,000


Avg. 99th PC values from 30 runs, at different sample sizes, for Claim # 18 (NOCTS 99th PC = 49.75)
Sample = 10000
Sample = 20000
Sample = 30000
Sample = 50000
Sample = 100000
Sample = 150000
Sample = 200000
Avg. 99th PC and 95% C.I.
50.0
49.8
49.6
49.4
49.2
Set 1
Set 2
Set 3
Set 4
Random Seeds Sets
Default Percentile
Alternative Percentile
95% C.I. for the 99th PC value is based on a sample size of 5,000,000


Avg. 99th PC values from 30 runs, at different sample sizes, for Claim # 19 (NOCTS 99th PC = 49.65)
Sample = 10000
Sample = 20000
Sample = 30000
Sample = 50000
Sample = 100000
Sample = 150000
Sample = 200000
Avg. 99th PC and 95% C.I.
50.2
50.0
49.8
49.6
49.4
Set 1
Set 2
Set 3
Set 4
Random Seeds Sets
Default Percentile
Alternative Percentile
95% C.I. for the 99th PC value is based on a sample size of 5,000,000


Avg. 99th PC values from 30 runs, at different sample sizes, for Claim # 21 (NOCTS 99th PC = 49.56)
Sample = 10000
Sample = 20000
Sample = 30000
Sample = 50000
Sample = 100000
Sample = 150000
Sample = 200000
Avg. 99th PC and 95% C.I.
50.0
49.8
49.6
49.4
Set 1
Set 2
Set 3
Set 4
Random Seeds Sets
Default Percentile
Alternative Percentile
95% C.I. for the 99th PC value is based on a sample size of 5,000,000


Avg. 99th PC values from 30 runs, at different sample sizes, for Claim # 22 (NOCTS 99th PC = 49.68)
Sample = 10000
Sample = 20000
Sample = 30000
Sample = 50000
Sample = 100000
Sample = 150000
Sample = 200000
Avg. 99th PC and 95% C.I.
50.50
50.25
50.00
49.75
49.50
49.25
Set 1
Set 2
Set 3
Set 4
Random Seeds Sets
Default Percentile
Alternative Percentile
95% C.I. for the 99th PC value is based on a sample size of 5,000,000


Avg. 99th PC values from 30 runs, at different sample sizes, for Claim # 24 (NOCTS 99th PC = 50.08)
Sample = 10000
Sample = 20000
Sample = 30000
Sample = 50000
Sample = 100000
Sample = 150000
Sample = 200000
Avg. 99th PC and 95% C.I.
50.25
50.00
49.75
49.50
49.25
49.00
Set 1
Set 2
Set 3
Set 4
Random Seeds Sets
Default Percentile
Alternative Percentile
95% C.I. for the 99th PC value is based on a sample size of 5,000,000

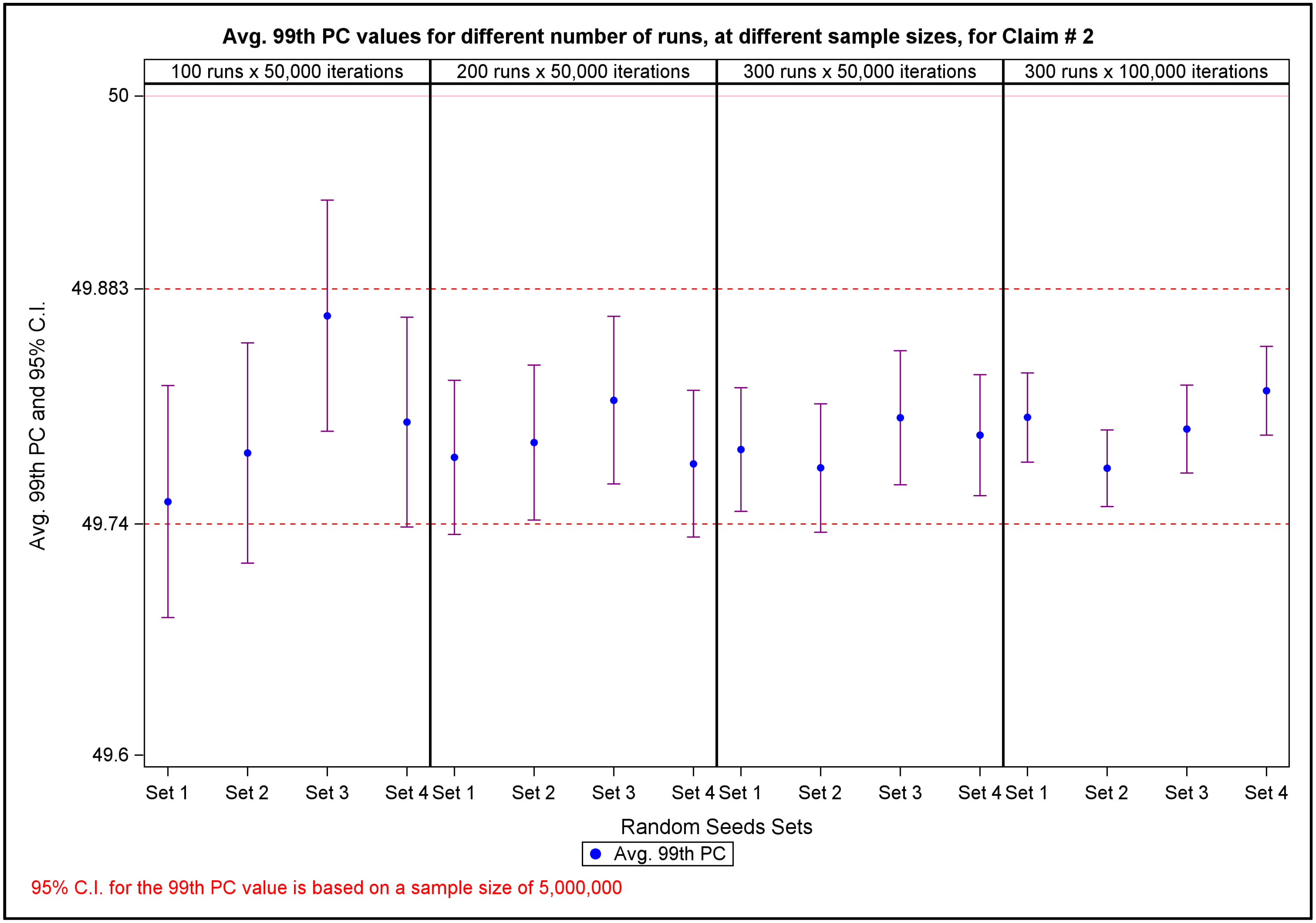
APPENDIX B

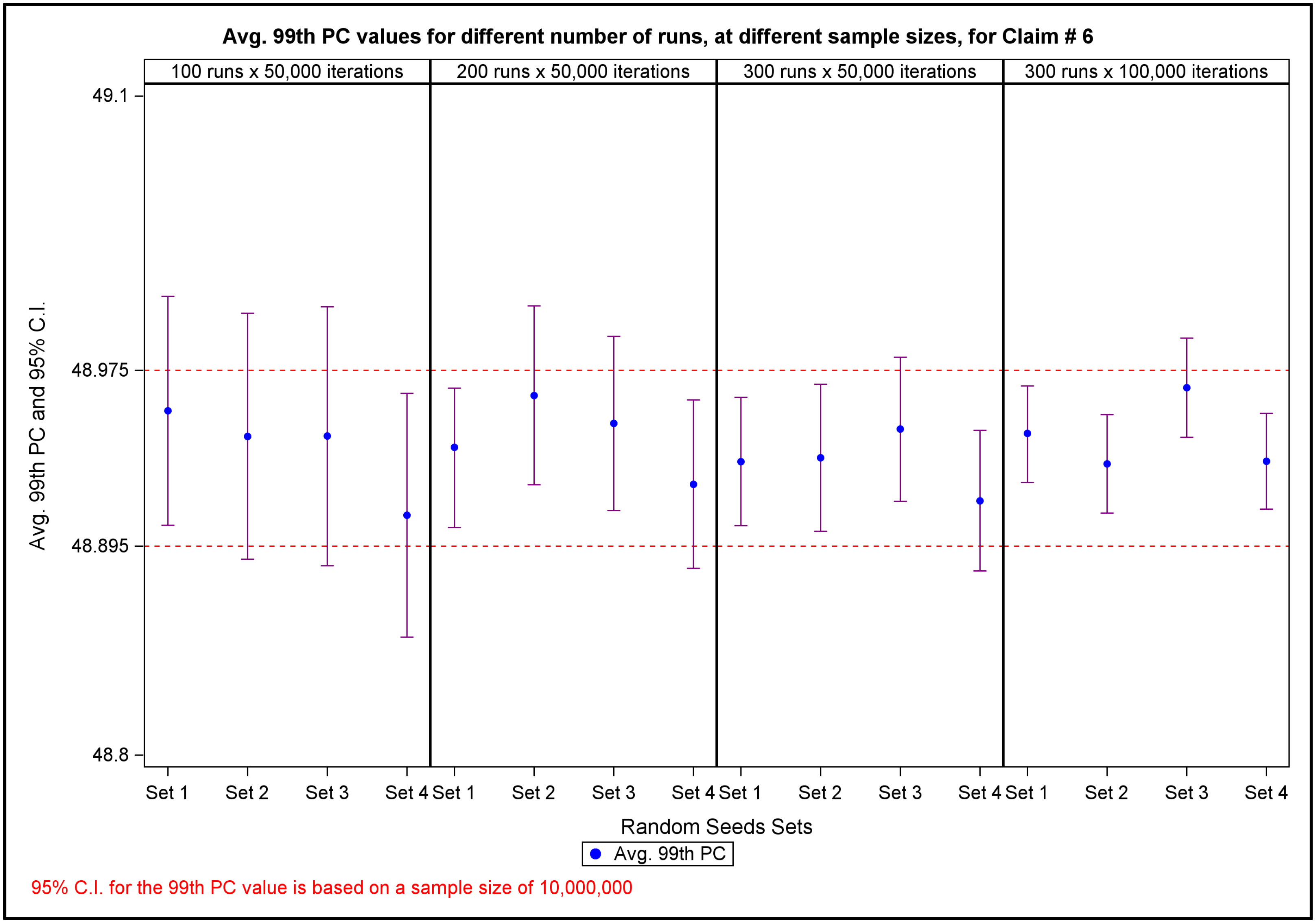
This appendix presents plots showing the convergence of the Avg. 99th PC values for 15 cases, obtained for the following four scenarios:

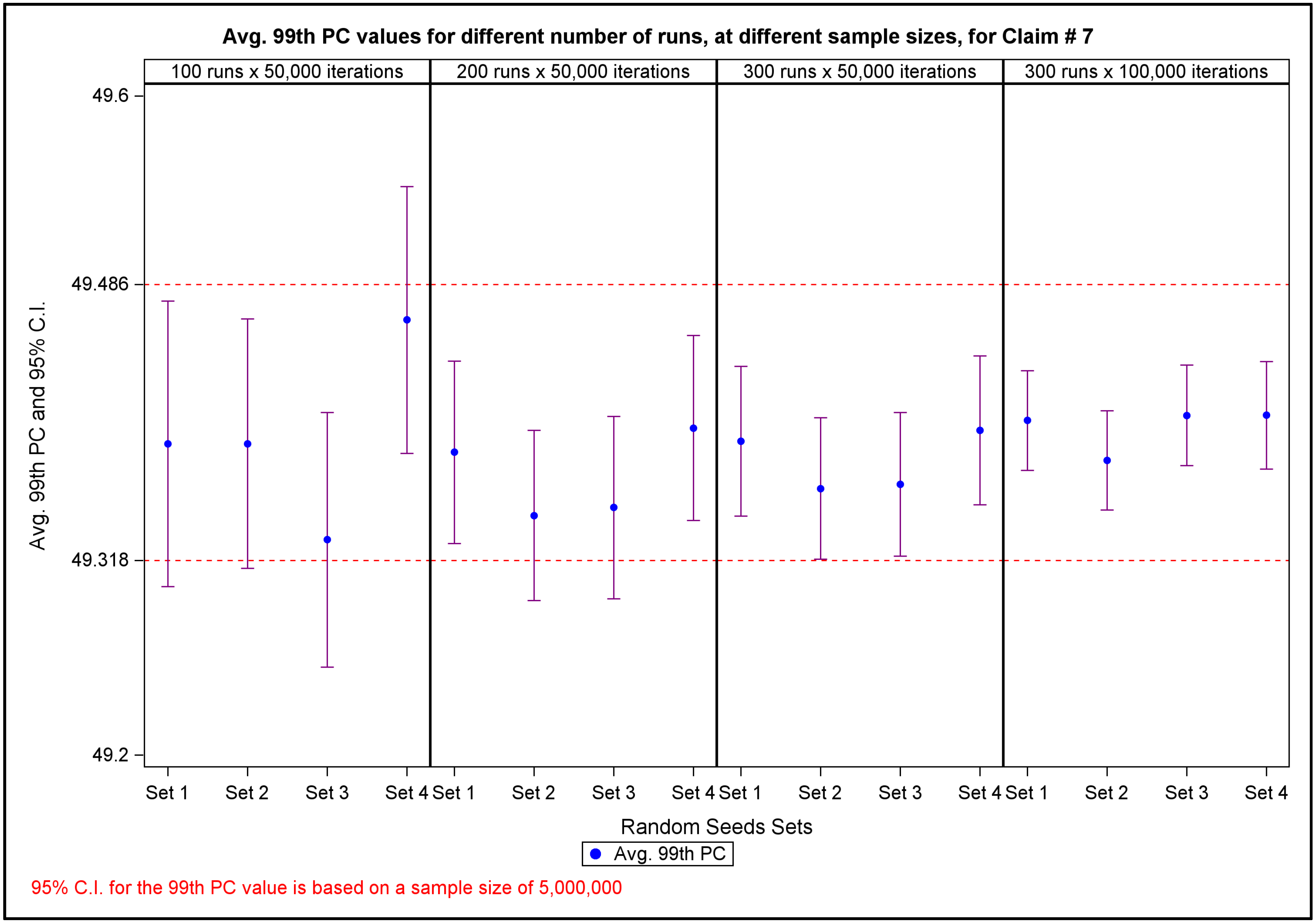
- Scenario 1: 100 runs with 50,000 iterations
- Scenario 2: 200 runs with 50,000 iterations
- Scenario 3: 300 runs with 50,000 iterations
- Scenario 4: 300 runs with 100,000 iterations.

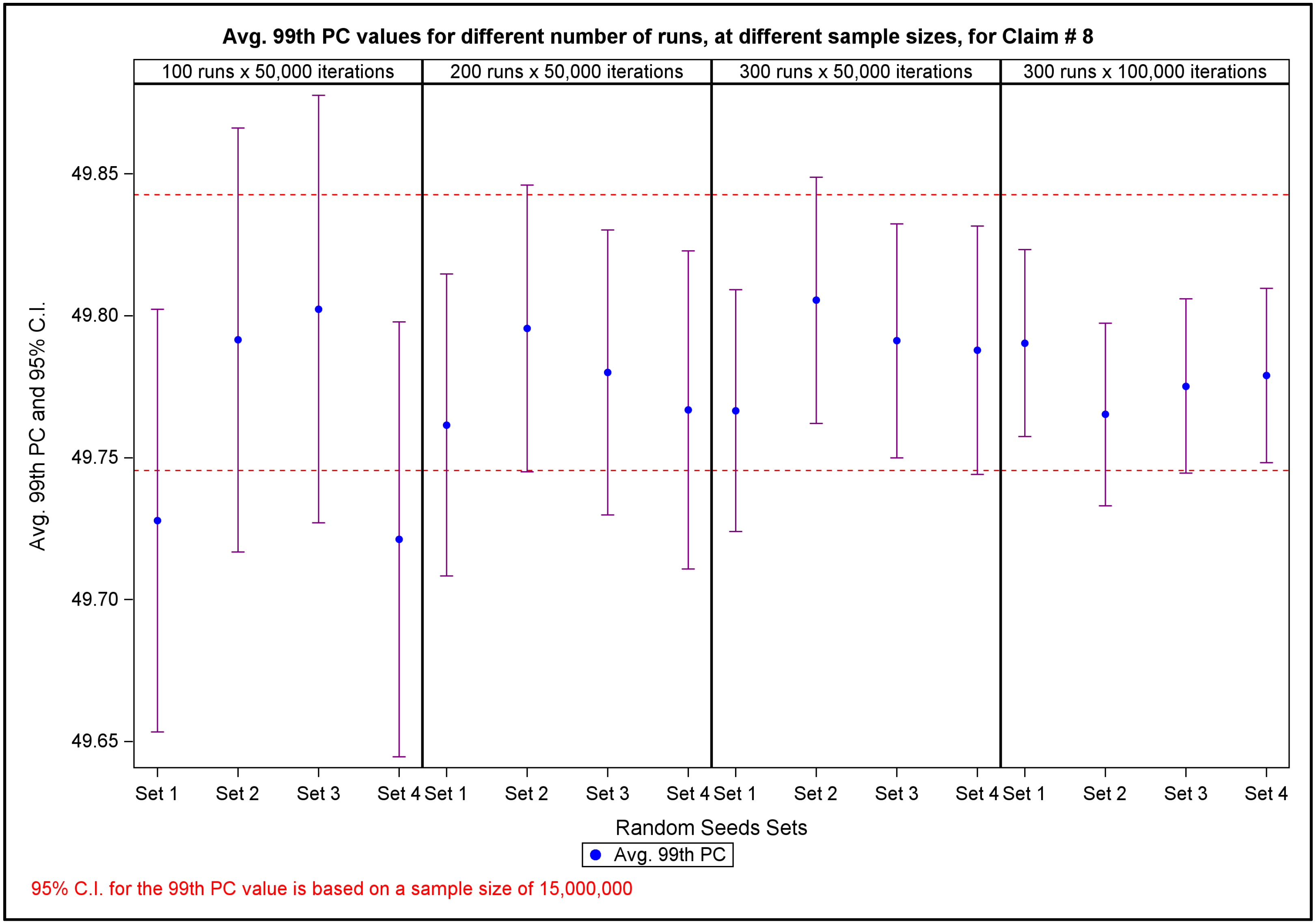
For each scenario, four sets of either 100, 200, or 300 random seeds were used (the same four sets of random seeds were used in each scenario).

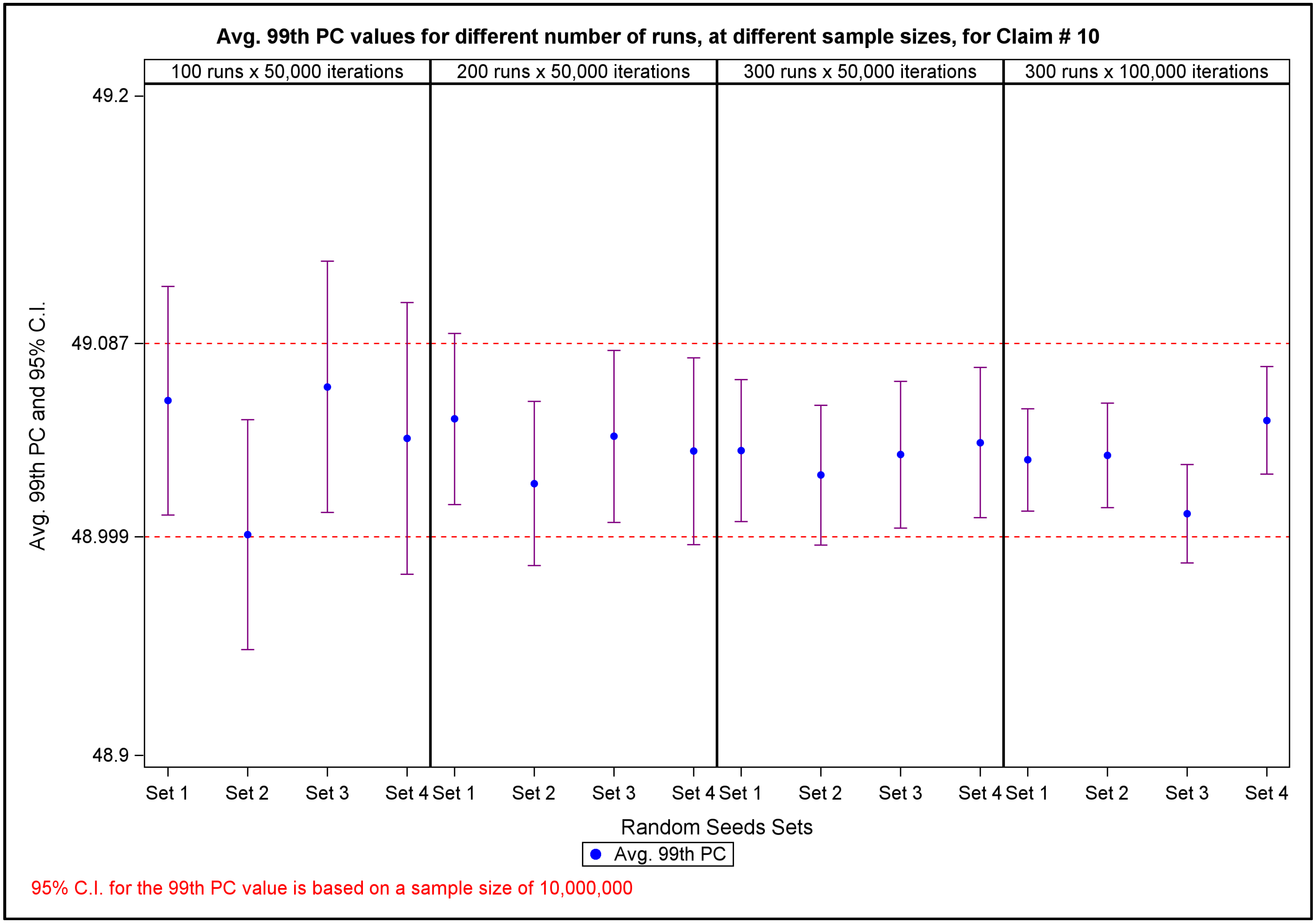
The 95% confidence interval for the theoretical 99th PC value for each case was obtained using a non-parametric method (Meeker 2017) either directly in the IREP version, or by combining multiple data sets with individual PC values obtained from IREP, and then using the non-parametric method for the combined data set with all the individual PC values.

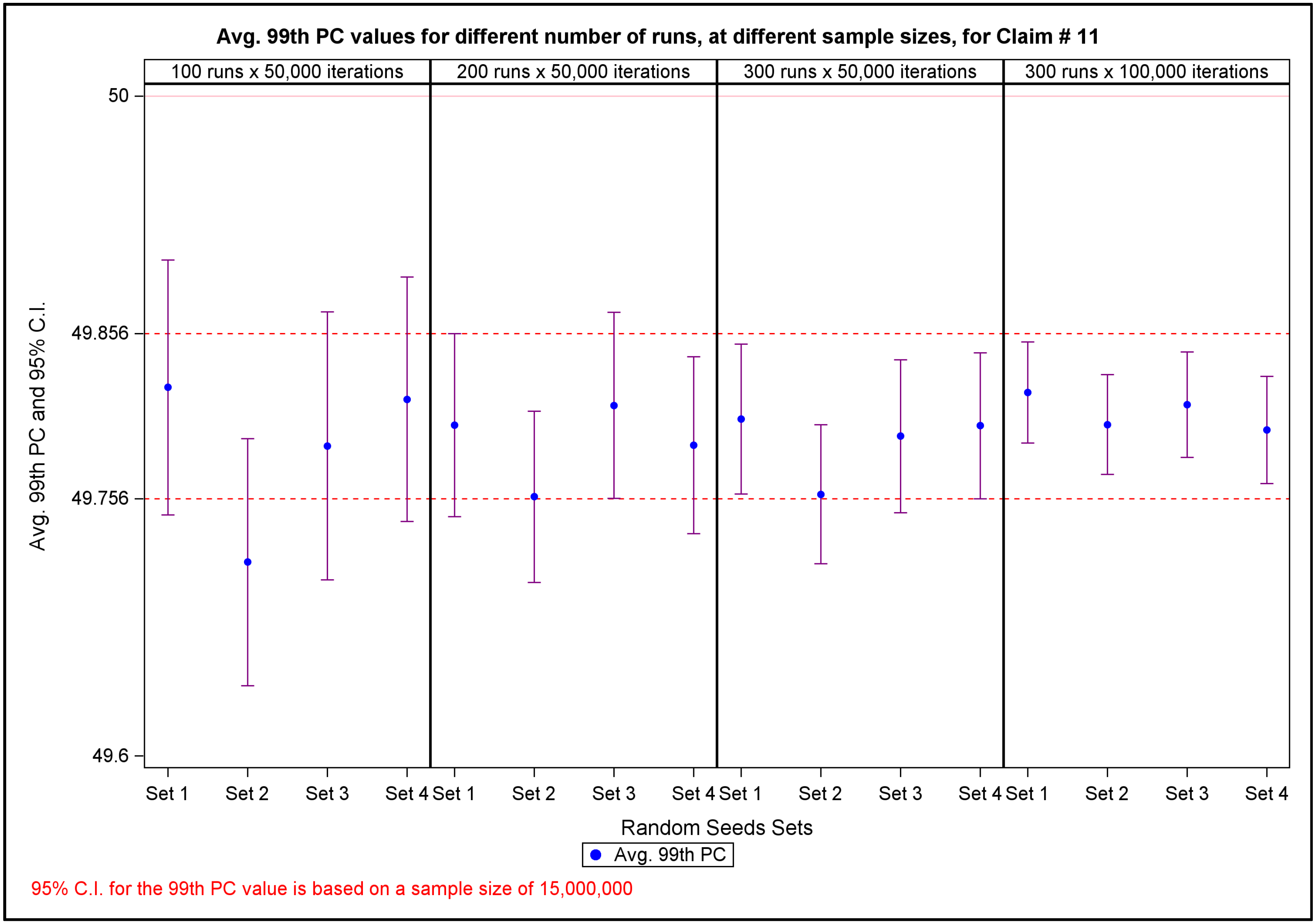

Avg. 99th PC values for different number of runs, at different sample sizes, for Claim # 2
100 runs x 50,000 iterations
200 runs x 50,000 iterations
300 runs x 50,000 iterations
300 runs x 100,000 iterations
50
49.883
49.74
49.6
Avg. 99th PC and 95% C.I.
Set 1
Set 2
Set 3
Set 4
Random Seeds Sets
Avg. 99th PC
95% C.I. for the 99th PC value is based on a sample size of 5,000,000

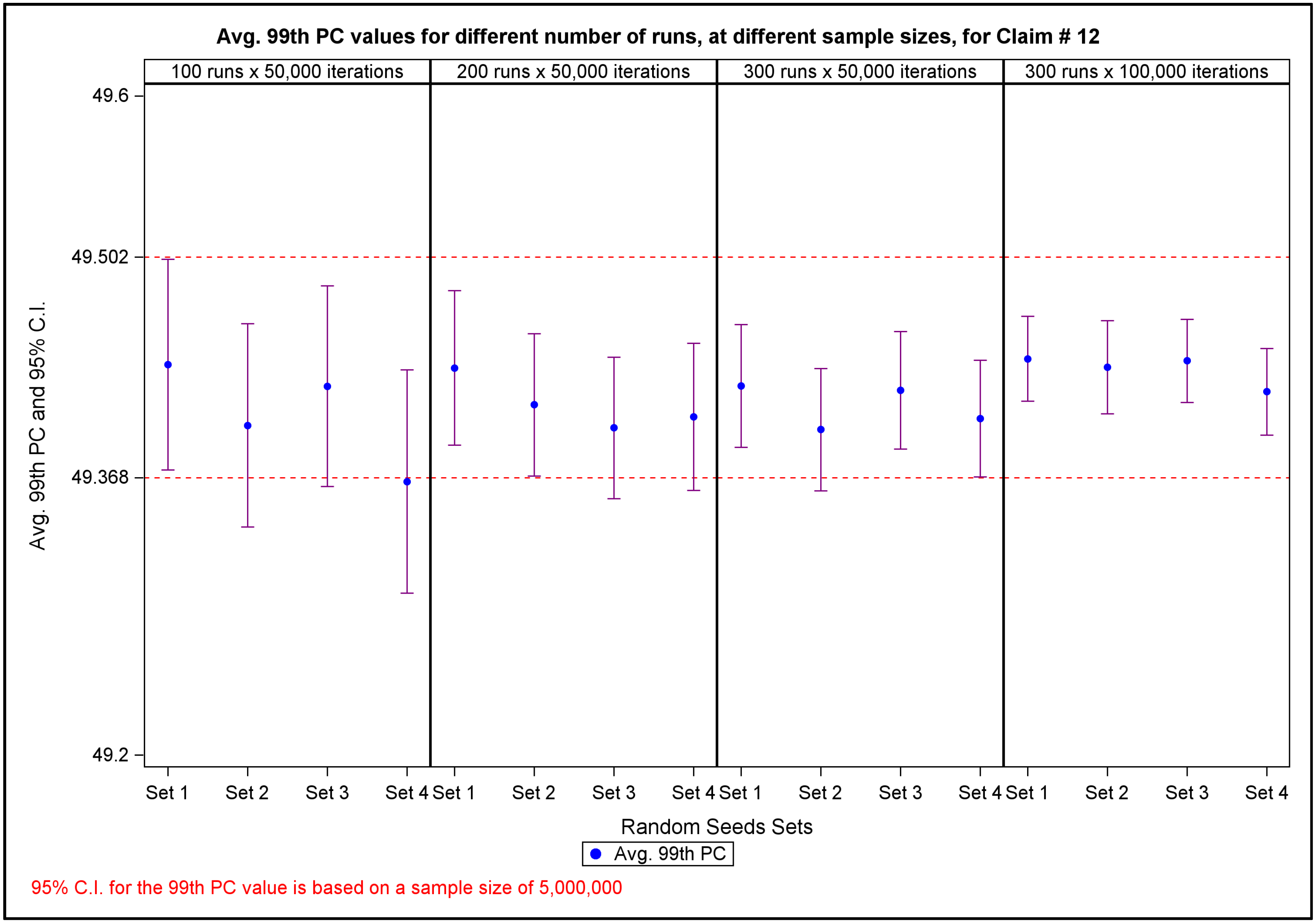

Avg. 99th PC values for different number of runs, at different sample sizes, for Claim # 6
100 runs x 50,000 iterations
200 runs x 50,000 iterations
300 runs x 50,000 iterations
300 runs x 100,000 iterations
Avg. 99th PC and 95% C.I.
49.1
48.975
48.895
48.8
Set 1
Set 2
Set 3
Set 4
Random Seeds Sets
Avg. 99th PC
95% C.I. for the 99th PC value is based on a sample size of 10,000,000

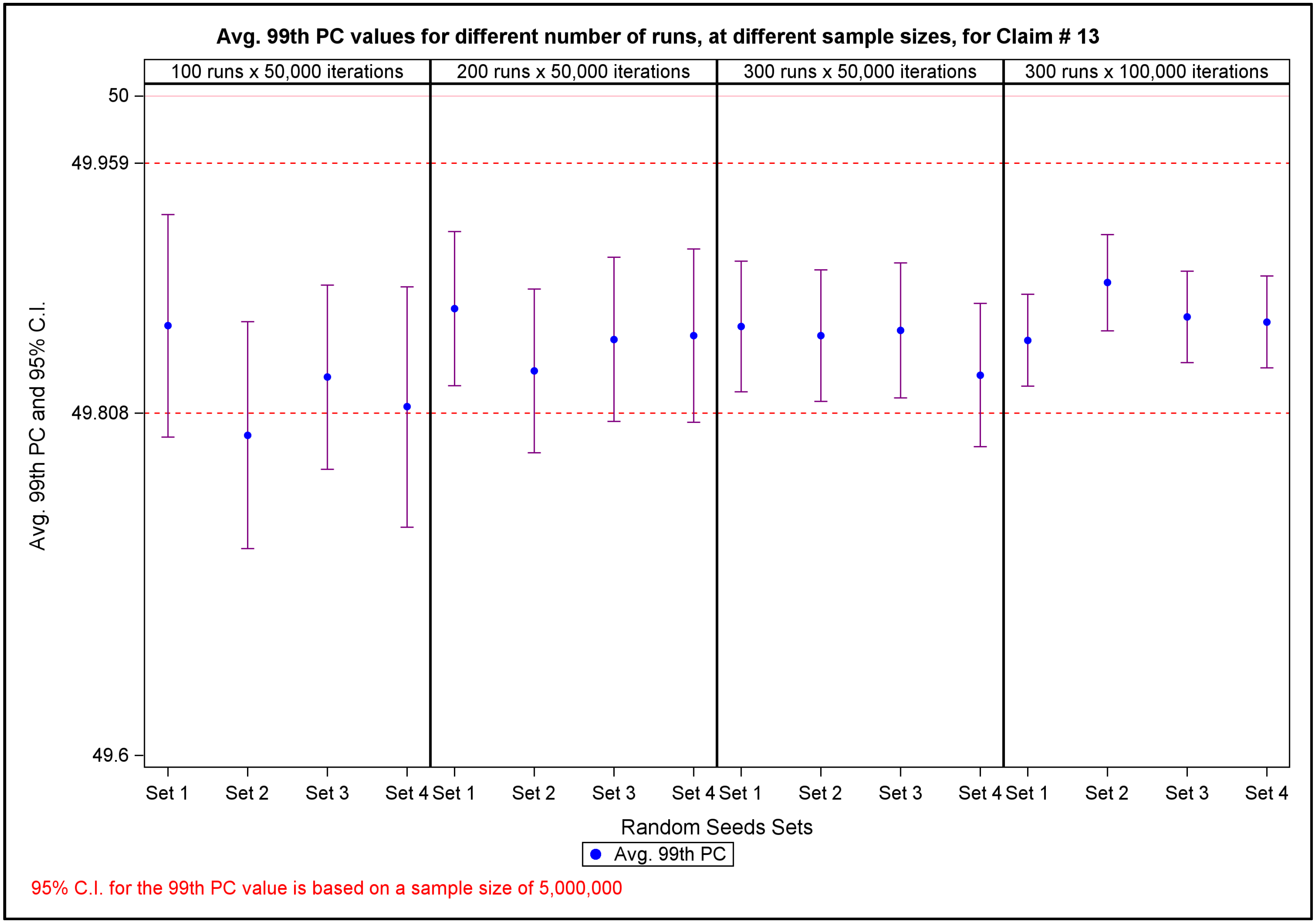

Avg. 99th PC values for different number of runs, at different sample sizes, for Claim # 7
100 runs x 50,000 iterations
200 runs x 50,000 iterations
300 runs x 50,000 iterations
300 runs x 100,000 iterations
Avg. 99th PC and 95% C.I.
49.6
49.486
49.318
49.2
Set 1
Set 2
Set 3
Set 4
Random Seeds Sets
Avg. 99th PC
95% C.I. for the 99th PC value is based on a sample size of 5,000,000

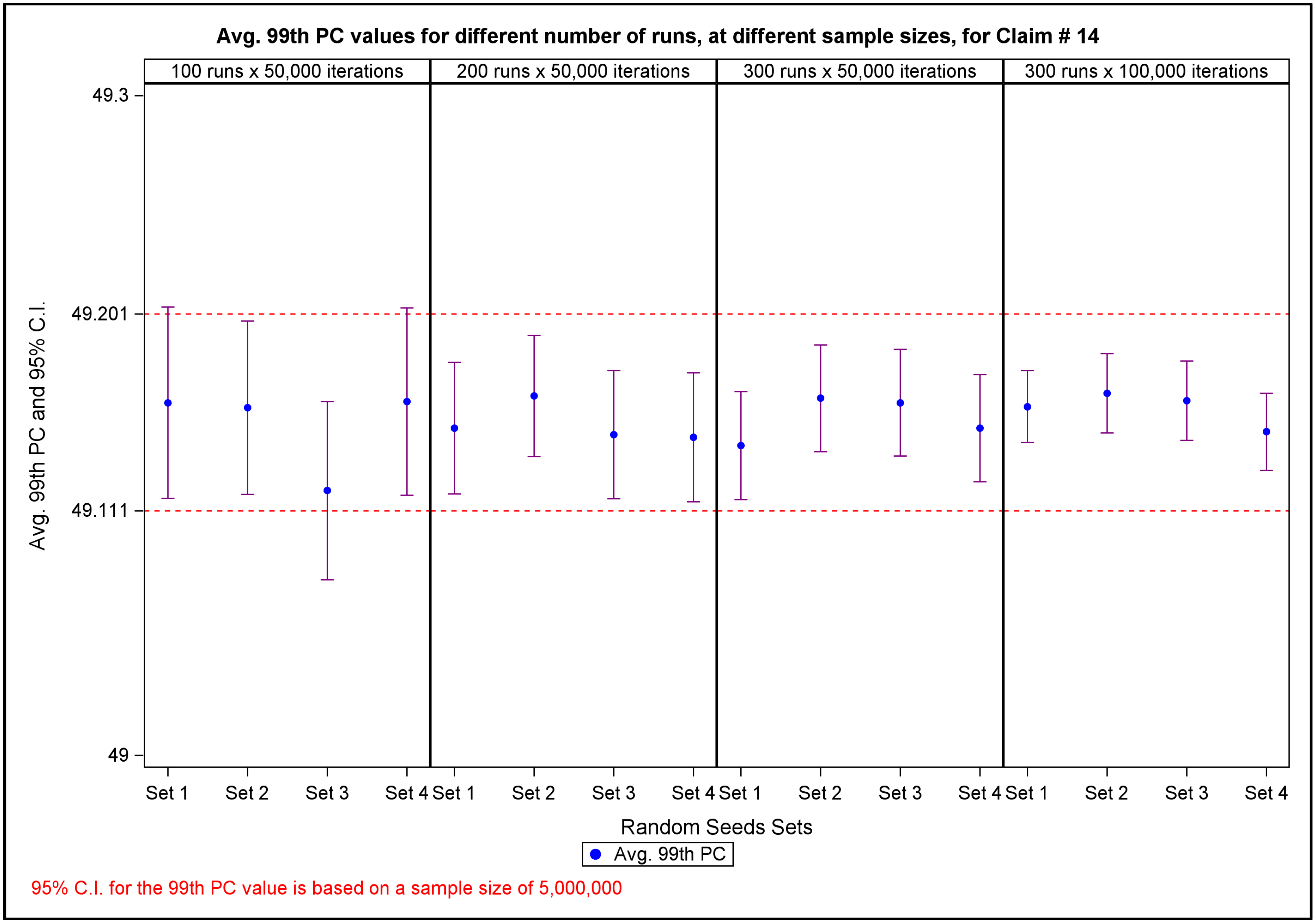

Avg. 99th PC values for different number of runs, at different sample sizes, for Claim # 8
100 runs x 50,000 iterations
200 runs x 50,000 iterations
300 runs x 50,000 iterations
300 runs x 100,000 iterations
Avg. 99th PC and 95% C.I.
49.85
49.80
49.75
49.70
49.65
Set 1
Set 2
Set 3
Set 4
Random Seeds Sets
Avg. 99th PC
95% C.I. for the 99th PC value is based on a sample size of 15,000,000

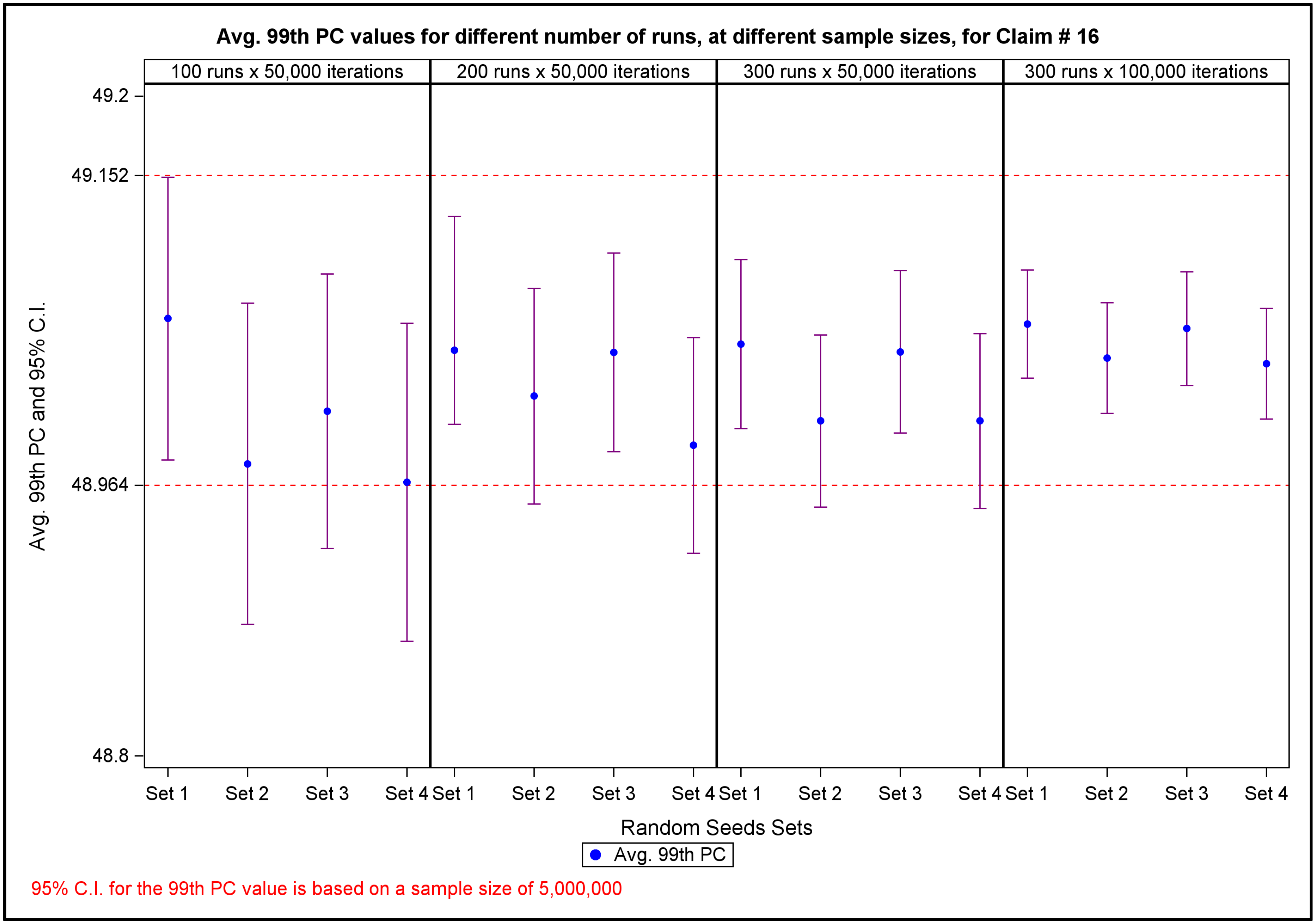

Avg. 99th PC values for different number of runs, at different sample sizes, for Claim # 10
100 runs x 50,000 iterations
200 runs x 50,000 iterations
300 runs x 50,000 iterations
300 runs x 100,000 iterations
49.2
49.087
48.999
48.9
Avg. 99th PC and 95% C.I.
Set 1
Set 2
Set 3
Set 4
Random Seeds Sets
Avg. 99th PC
95% C.I. for the 99th PC value is based on a sample size of 10,000,000

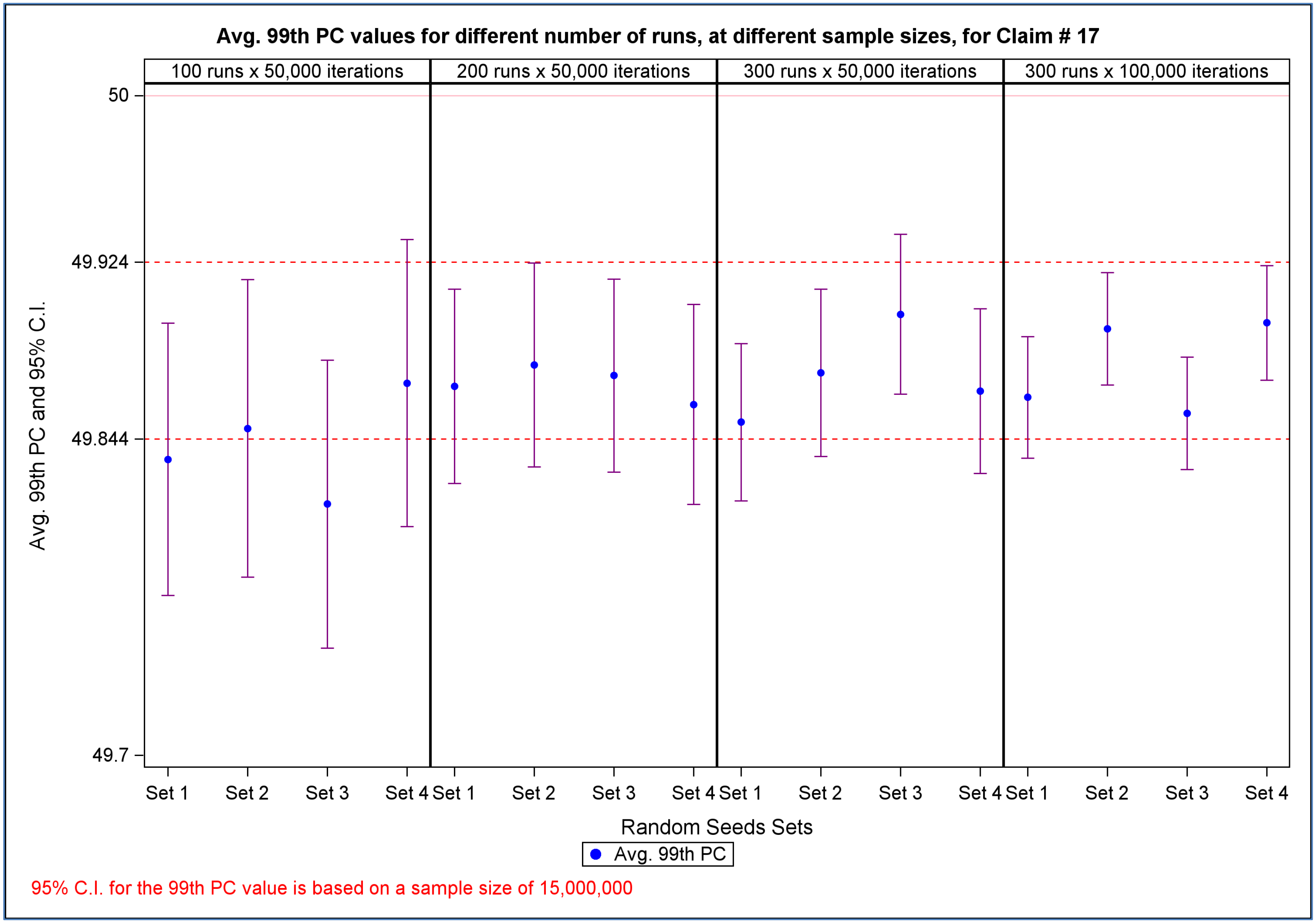

Avg. 99th PC values for different number of runs, at different sample sizes, for Claim # 11
100 runs x 50,000 iterations
200 runs x 50,000 iterations
300 runs x 50,000 iterations
300 runs x 100,000 iterations
50
49.856
49.756
49.6
Avg. 99th PC and 95% C.I.
Set 1
Set 2
Set 3
Set 4
Random Seeds Sets
Avg. 99th PC
95% C.I. for the 99th PC value is based on a sample size of 15,000,000

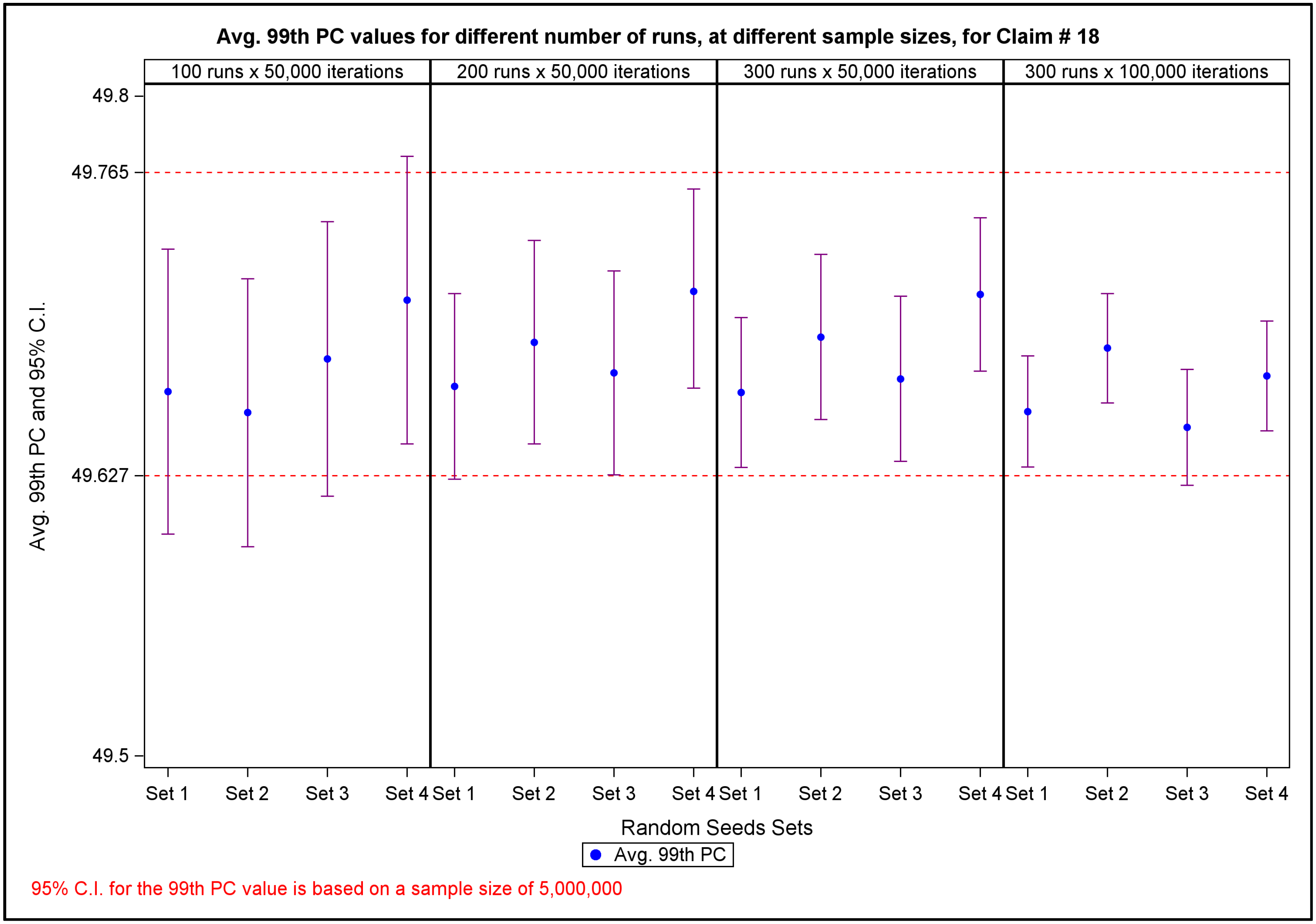

Avg. 99th PC values for different number of runs, at different sample sizes, for Claim # 12
100 runs x 50,000 iterations
200 runs x 50,000 iterations
300 runs x 50,000 iterations
300 runs x 100,000 iterations
Avg. 99th PC and 95% C.I.
49.6
49.502
49.368
49.2
Set 1
Set 2
Set 3
Set 4
Random Seeds Sets
Avg. 99th PC
95% C.I. for the 99th PC value is based on a sample size of 5,000,000

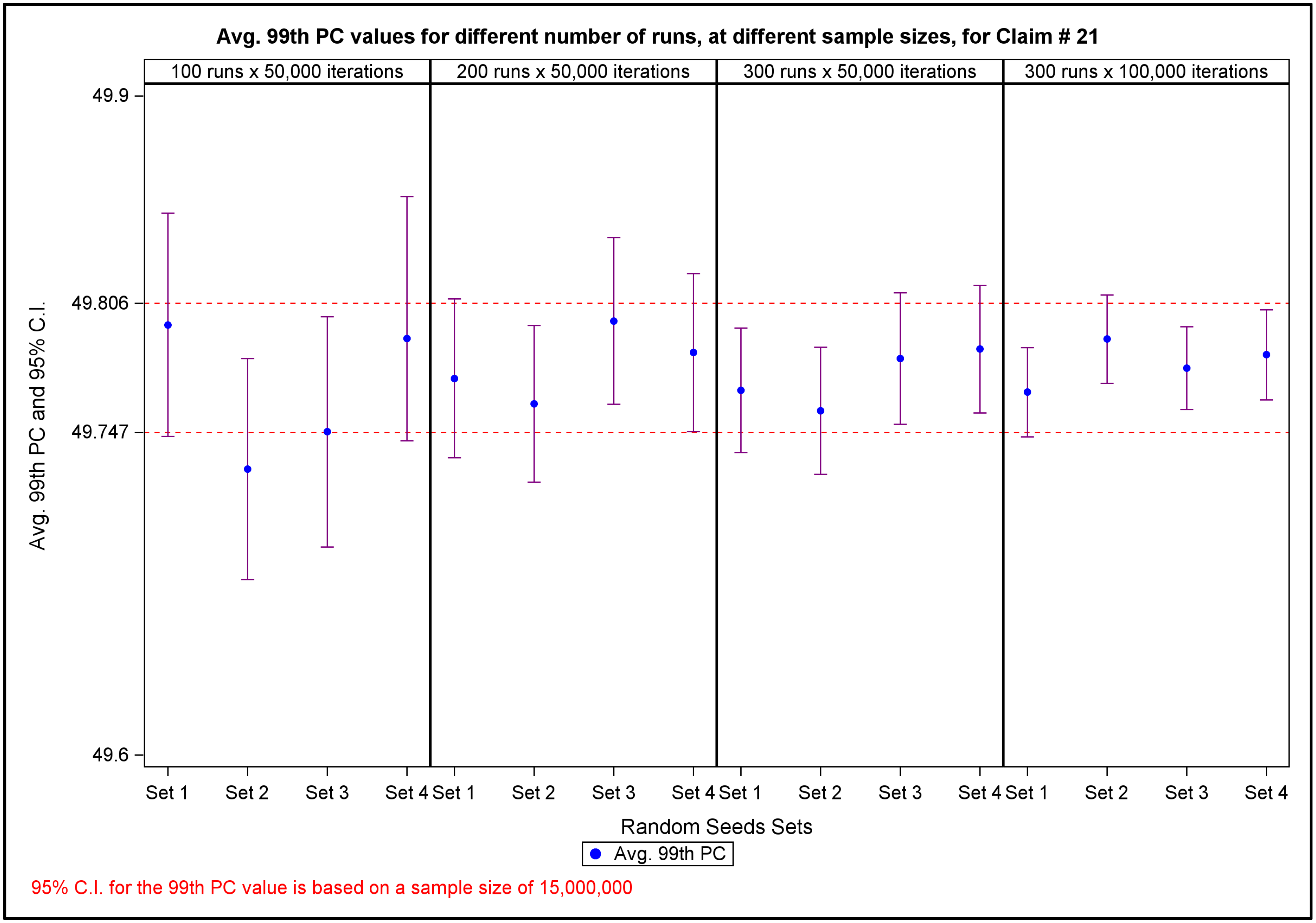

Avg. 99th PC values for different number of runs, at different sample sizes, for Claim # 13
100 runs x 50,000 iterations
200 runs x 50,000 iterations
300 runs x 50,000 iterations
300 runs x 100,000 iterations
Avg. 99th PC and 95% C.I.
50
49.959
49.808
49.6
Set 1
Set 2
Set 3
Set 4
Random Seeds Sets
Avg. 99th PC
95% C.I. for the 99th PC value is based on a sample size of 5,000,000

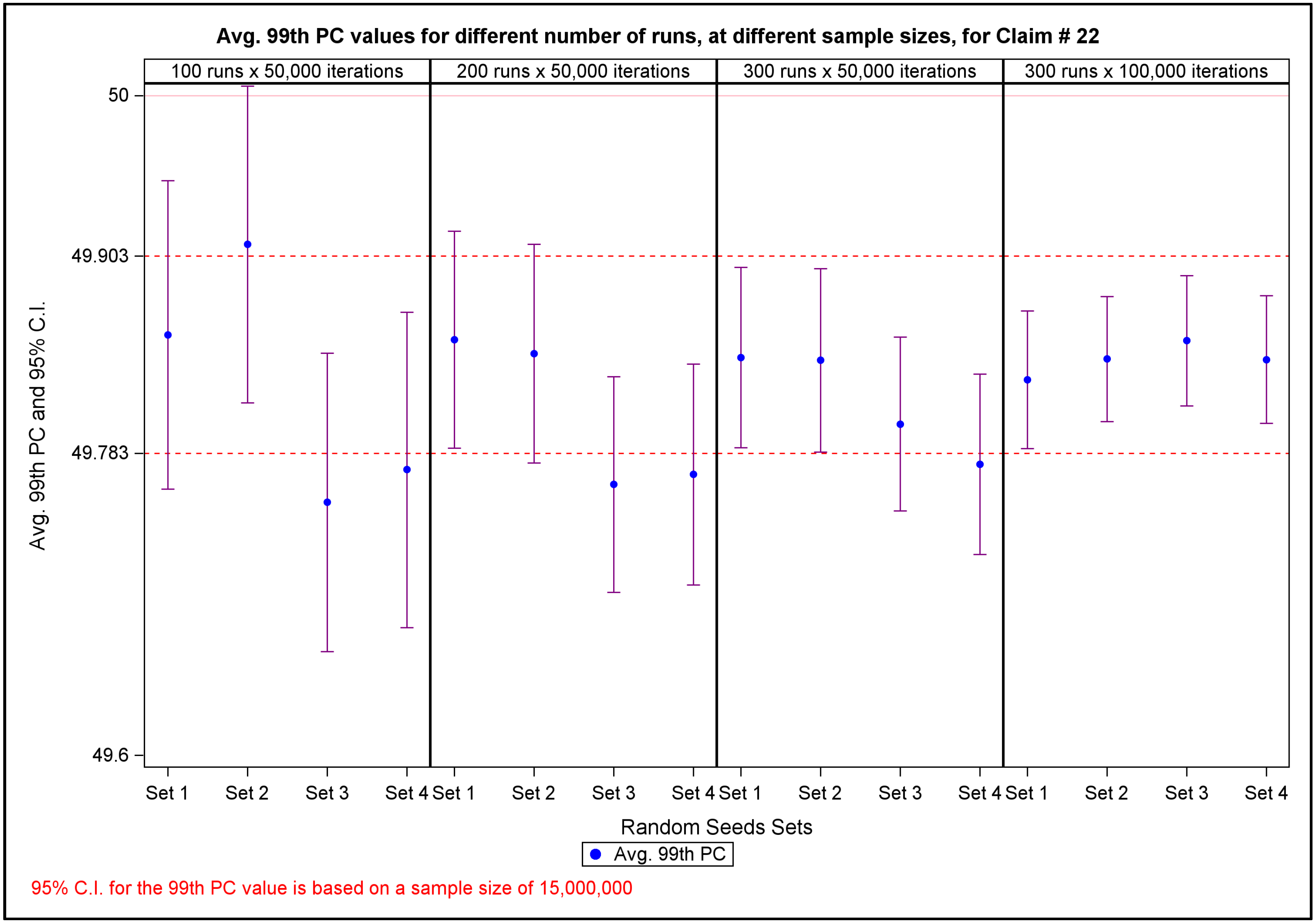

Avg. 99th PC values for different number of runs, at different sample sizes, for Claim # 14
100 runs x 50,000 iterations
200 runs x 50,000 iterations
300 runs x 50,000 iterations
300 runs x 100,000 iterations
Avg. 99th PC and 95% C.I.
49.3
49.201
49.111
49
Set 1
Set 2
Set 3
Set 4
Random Seeds Sets
Avg. 99th PC
95% C.I. for the 99th PC value is based on a sample size of 5,000,000

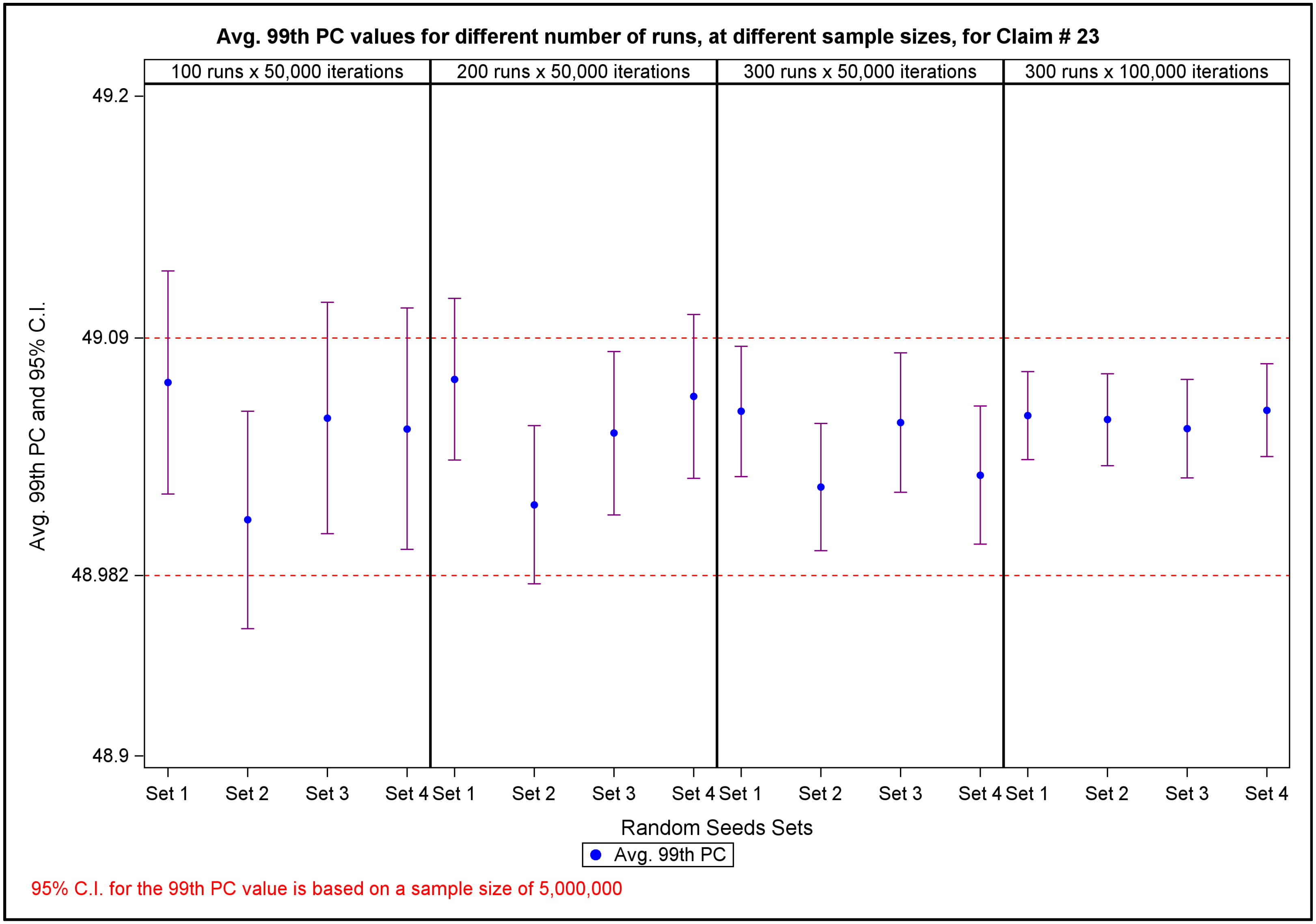

Avg. 99th PC values for different number of runs, at different sample sizes, for Claim # 16
100 runs x 50,000 iterations
200 runs x 50,000 iterations
300 runs x 50,000 iterations
300 runs x 100,000 iterations
Avg. 99th PC and 95% C.I.
49.2
49.152
48.964
48.8
Set 1
Set 2
Set 3
Set 4
Random Seeds Sets
Avg. 99th PC
95% C.I. for the 99th PC value is based on a sample size of 5,000,000


Avg. 99th PC values for different number of runs, at different sample sizes, for Claim # 17
100 runs x 50,000 iterations
200 runs x 50,000 iterations
300 runs x 50,000 iterations
300 runs x 100,000 iterations
50
49.924
49.844
49.7
Avg. 99th PC and 95% C.I.
Set 1
Set 2
Set 3
Set 4
Random Seeds Sets
Avg. 99th PC
95% C.I. for the 99th PC value is based on a sample size of 15,000,000


Avg. 99th PC values for different number of runs, at different sample sizes, for Claim # 18
100 runs x 50,000 iterations
200 runs x 50,000 iterations
300 runs x 50,000 iterations
300 runs x 100,000 iterations
Avg. 99th PC and 95% C.I.
49.8
49.765
49.627
49.5
Set 1
Set 2
Set 3
Set 4
Random Seeds Sets
Avg. 99th PC
95% C.I. for the 99th PC value is based on a sample size of 5,000,000


Avg. 99th PC values for different number of runs, at different sample sizes, for Claim # 21
100 runs x 50,000 iterations
200 runs x 50,000 iterations
300 runs x 50,000 iterations
300 runs x 100,000 iterations
Avg. 99th PC and 95% C.I.
49.9
49.806
49.747
49.6
Set 1
Set 2
Set 3
Set 4
Random Seeds Sets
Avg. 99th PC
95% C.I. for the 99th PC value is based on a sample size of 15,000,000


Avg. 99th PC values for different number of runs, at different sample sizes, for Claim # 22
100 runs x 50,000 iterations
200 runs x 50,000 iterations
300 runs x 50,000 iterations
300 runs x 100,000 iterations
50
49.903
49.783
49.6
Avg. 99th PC and 95% C.I.
Set 1
Set 2
Set 3
Set 4
Random Seeds Sets
Avg. 99th PC
95% C.I. for the 99th PC value is based on a sample size of 15,000,000


Avg. 99th PC values for different number of runs, at different sample sizes, for Claim # 23
100 runs x 50,000 iterations
200 runs x 50,000 iterations
300 runs x 50,000 iterations
300 runs x 100,000 iterations
Avg. 99th PC and 95% C.I.
49.2
49.09
48.982
48.9
Set 1
Set 2
Set 3
Set 4
Random Seeds Sets
Avg. 99th PC
95% C.I. for the 99th PC value is based on a sample size of 5,000,000