Increasing the Accuracy of Probability of Causation

A Summary of Research

White Paper

National Institute for Occupational Safety and Health (NIOSH)

July 2023

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. The 99^{th} percentile of that distribution is used in decision making. Currently, 2,000 samples are used in IREP to determine 99^{th} 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) to improve precision and the average of the resulting thirty 99^{th} percentiles (Avg. 99^{th} PC value) is used for compensation. However, even at 30 runs of 10,000 iterations each, the Avg. 99^{th} PC value can vary with changing random seeds by up to ±0.5, a variability that is significant when 99^{th} percentiles of PC are close to 50%.

This report summarizes a series of studies dedicated to further increasing precision in the estimated 99^{th} percentile of PC, when expected values are close to the threshold of 50%.

As a first step, ten commonly used methods to estimate a percentile (e.g., 99^{th}) of a pdf described by multiple samples were identified and tested on hypothetical cases and real claims. The performance of a percentile method differs, depending on the percentile level, sample size, type of underlying distribution, and magnitude of uncertainty and of positive skewness, but, in general, the larger the sample size, the smaller the difference between the values of the same percentile produced by different methods. For all tested scenarios, the sample percentile method used by default in the SAS software had the smallest relative bias and smallest relative root mean squared error at all the sample sizes, when estimating the 99^{th} percentile. The sample percentile method currently used in NIOSH-IREP tends to underestimate the theoretical 99^{th} percentile in all the scenarios tested, especially for sample sizes lower than 20,000 values. For sample sizes greater than 20,000 the difference between methods becomes small enough and can be neglected. Thus, it is preferrable to use 20,000 or more samples in PC calculations using NIOSH-IREP for EEIOCPA claim evaluations.

Secondly, the convergence of the 99^{th} PC values as well as of the Avg. 99^{th} PC values, as the number of iterations increases was shown for multiple cases with PC values close to 50%, and for each of these cases there is a clear trend for both quantities to converge to a stable value that represents the theoretical 99^{th} percentile from the distribution of PC values for that case.

The theoretical 99^{th} PC value of the PC probability distribution for a case cannot be determined exactly, but a 95% confidence interval for the estimated 99^{th} 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 99^{th} PC value will be, where the length is defined as the upper bound minus the lower bound of the 95% confidence interval. For a large enough sample size, 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. 99^{th} 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. 99^{th} PC values lower than 0.1 were obtained for all tested cases when using 300 runs at 100,000 iterations.

To increase efficiency and minimize running time, a Predictive Tool was developed to determine the minimum number of iterations needed to achieve a desired precision measured by the length of the 95% confidence interval of the Avg. 99^{th} PC values. For a 0.15 precision level, the number of iterations needed for the 23 test cases with Avg. 99^{th} 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. 99^{th} PC values can be efficiently and reliably reduced below one of the desirable precision levels of 0.15 or 0.10.

Based on the analyses described in this report, the following actions are recommended regarding the settings that need to be changed in the current NIOSH-IREP versions, as well as related to how cases are selected to be run on the two IREP versions.

1. The number of iterations used by default in the current IREP v.5.9 version should be increased from 2,000 iterations to 20,000 iterations.

2. The number of iterations used by default in the current IREP EE v.5.9 version should be increased from 10,000 iterations to 20,000 iterations.

3. The range for the cases that are selected to be run on the IREP EE version can be modified from the current range of (45.0%, 51.99%) to a range of (47.0%, 53.0%).

4. Incorporate the IREP Predictive Tool into the IREP EE version and change IREP EE to run cases with either 30 random seeds or 300 random seeds. The new IREP EE version would feel and operate just like the current IREP EE version, with no discernible difference for a user.

   The IREP Predictive Tool implemented in IREP EE should be used for cases with Avg. 99^{th} PC values obtained from the IREP EE version in the range of (49.5%, 50.5%), to determine the minimum number of iterations required to achieve a precision of less than or equal to 0.10 around the Avg. 99^{th} PC value. Those cases will then be rerun on IREP EE with 300 runs and the number of iterations suggested by the IREP Predictive Tool.

TABLE OF CONTENTS

1.0 INTRODUCTION

This paper is a summary of the main findings from three white papers: “Comparison of Several Percentile Definitions” (Stancescu et al. 2019), “Effect of Alternative Percentile Definition on PC Values” (Stancescu et al. 2021), and “Increasing the Accuracy of the 99th Percentile of PC” (Stancescu et al. 2022). The document summarizes the comparison of several percentile definitions, the effect of Alternative Percentile Definitions on the PC values, the PC variability at different number of iterations in the NIOSH-IREP software, the convergence of the 99th PC values as the number of iterations is increased in IREP, the different settings that can used to achieve a desired precision for the 99th PC values, the use of the IREP Predictive Tool to minimize the number of iterations, and the validation of the 99th PC values from the IREP and IREP EE versions.

A list of recommendations is provided at the end of the document regarding the settings that need to be changed in the current NIOSH-IREP v.5.9 versions, as well as related to how the cases are selected to be run on the two IREP versions.

2.0 COMPARISON OF SEVERAL PERCENTILE DEFINITIONS

There is no standard method to estimate a percentile (quantile) from multiple samples obtained from a probability distribution. Multiple definitions are currently in use, with each software package using different methods to compute the percentiles. All the probabilistic modeling and risk analysis software packages like Crystal Ball (Oracle 2010), @Risk (Palisade 2016), Analytica (Lumina 2015), and Model Risk (Vose 2012) have different methods implemented to compute the percentiles. Statistical software packages like SAS or R have multiple definitions implemented to compute a percentile, with one method being used by default. There are several articles in statistical literature dealing with the issue of which percentile definition to use. Hyndman and Fan (1996) evaluated nine different methods for computing the quantiles relative to six desirable properties and advised standardizing the quantile definition in statistical software.

In the absence of a standard method to compute a percentile, there is a need to clarify how the 99th percentile is estimated in the NIOSH-IREP software, and how accurate this estimation is during the process of the Monte Carlo computations used for the purpose of estimating the Probability of Causation (PC) values.

Hyndman and Fan (1996) analyzed nine different methods of computing a sample’s quantiles. Their goal was to advocate for a standardized definition of percentiles so that percentiles reported by different statistical software will be computed according to the same formula. While this goal hasn’t been achieved yet, the Hyndman and Fan (1996) paper is the most widely cited article regarding this issue, and most of the statistical software use one of the nine methods described by Hyndman and Fan (1996). These nine methods are presented here, and some of the software packages that use those methods as default are identified. Often software packages have more than one method implemented, with one method being used by default, unless the user specifically selects a different method (Analytica has only one of these methods implemented, current versions of Excel have 2 of these methods implemented, SAS has 5 of methods implemented, and R has available all 9 methods).

Table 1 shows a summary of the nine sample quantile definitions from Hyndman and Fan (1996), reformatted from Wikipedia. In Table 1, N represents the number of samples (values) in the data set, $x_1, \ldots, x_N$ represent the sorted data values, and p is the percentile level. The nearest-rank method is the most common definition of a percentile, as introduced in elementary statistics, and determines the percentile as the value from the data that splits the sorted data set into two parts, with the lower part contains p percent of the data, and the upper part contains the rest of the data, i.e., $(100 - p)$ percent of the data. The first three methods estimate the percentiles using a discrete rounding scheme, while the other six methods use a continuous interpolation scheme.

All the methods used for computing the percentiles are non-parametric, which means that no assumption is made about the distribution for the population where the samples are taken.

Table 1. Summary of the nine sample quantile definitions, from Hyndman and Fan (1996)[a].

Method	h	Q_p	Description
Type 1 (nearest-rank)	$Np + 1/2$	$x_{\lceil h-1/2 \rceil}$	Inverse of the empirical cumulative distribution function.
Type 2 (default in SAS)	$Np + 1/2$	$(x_{\lceil h-1/2 \rceil} + x_{\lfloor h+1/2 \rfloor}) / 2$	Same as Type 1, but with averaging at discontinuities.
Type 3	Np	$x_{[h]}$	The observation numbered closest to Np.
Type 4 (@Risk, Crystal Ball)	Np	$x_{\lfloor h \rfloor} + (h - \lfloor h \rfloor)(x_{\lfloor h \rfloor+1} - x_{\lfloor h \rfloor})$	Linear interpolation of the empirical distribution function.
Type 5	$Np + 1/2$	$x_{\lfloor h \rfloor} + (h - \lfloor h \rfloor)(x_{\lfloor h \rfloor+1} - x_{\lfloor h \rfloor})$	Piecewise linear function where the knots are the values midway through the steps of the empirical distribution function.
Type 6 (Excel PERCENTILE.EXC)	$(N + 1)p$	$x_{\lfloor h \rfloor} + (h - \lfloor h \rfloor)(x_{\lfloor h \rfloor+1} - x_{\lfloor h \rfloor})$	Linear interpolation of the expectations for the order statistics for the uniform distribution on [0,1].
Type 7 (default in R, Analytica, Excel PERCENTILE)	$(N - 1)p + 1$	$x_{\lfloor h \rfloor} + (h - \lfloor h \rfloor)(x_{\lfloor h \rfloor+1} - x_{\lfloor h \rfloor})$	Linear interpolation of the modes for the order statistics for the uniform distribution on [0,1].
Type 8	$(N + 1/3)p + 1/3$	$x_{\lfloor h \rfloor} + (h - \lfloor h \rfloor)(x_{\lfloor h \rfloor+1} - x_{\lfloor h \rfloor})$	Linear interpolation of the approximate medians for order statistics.
Type 9	$(N + 1/4)p + 3/8$	$x_{\lfloor h \rfloor} + (h - \lfloor h \rfloor)(x_{\lfloor h \rfloor+1} - x_{\lfloor h \rfloor})$	The resulting quantile estimates are approximately unbiased for the expected order statistics if data is normally distributed.

[a] See also: https://en.wikipedia.org/wiki/Quantile

Figure 1 presents six of the nine methods from Table 1, plus the method used in the Vose software, for a small data set of 10 observations: {1, 2, 2, 3, 3, 4, 5, 5, 5, 9}. The 90th percentile computed using the seven methods ranges anywhere between the last two values in the data set, which are 5 and 9. The nearest-rank method assigns the lowest of the two values ($Q_{0.9} = 5$), while the Vose software assigns the highest of the two values ($Q_{0.9} = 9$). The default method in SAS (Type 2 method) assigns the middle of the two values ($Q_{0.9} = 7$), while the default method in R, which is also used by the Analytica software (Type 7 method), assigns the value ($Q_{0.9} = 5.4$) as the 90th percentile. Other values assigned by the other methods for the 90th percentile can vary anywhere between the last two values in the data set, as it can be seen in Figure 1 ($Q_{0.9} = 8.6$ for Type 6 method, $Q_{0.9} = 7.53$ for Type 8 method, and $Q_{0.9} = 7.4$ for Type 9 method). Not all the methods are displayed here, since some of them produce identical values for the percentiles; for the chosen data set and the 90th percentile, the Type 3 and Type 4 methods produce the same result as the Type 1 method, and the Type 5 method produce the same result as Type 2 method. From the nine methods used in this example, three methods (Type 1, Type 2, and Vose) use a discrete rounding scheme, while the other methods use a continuous interpolation scheme.

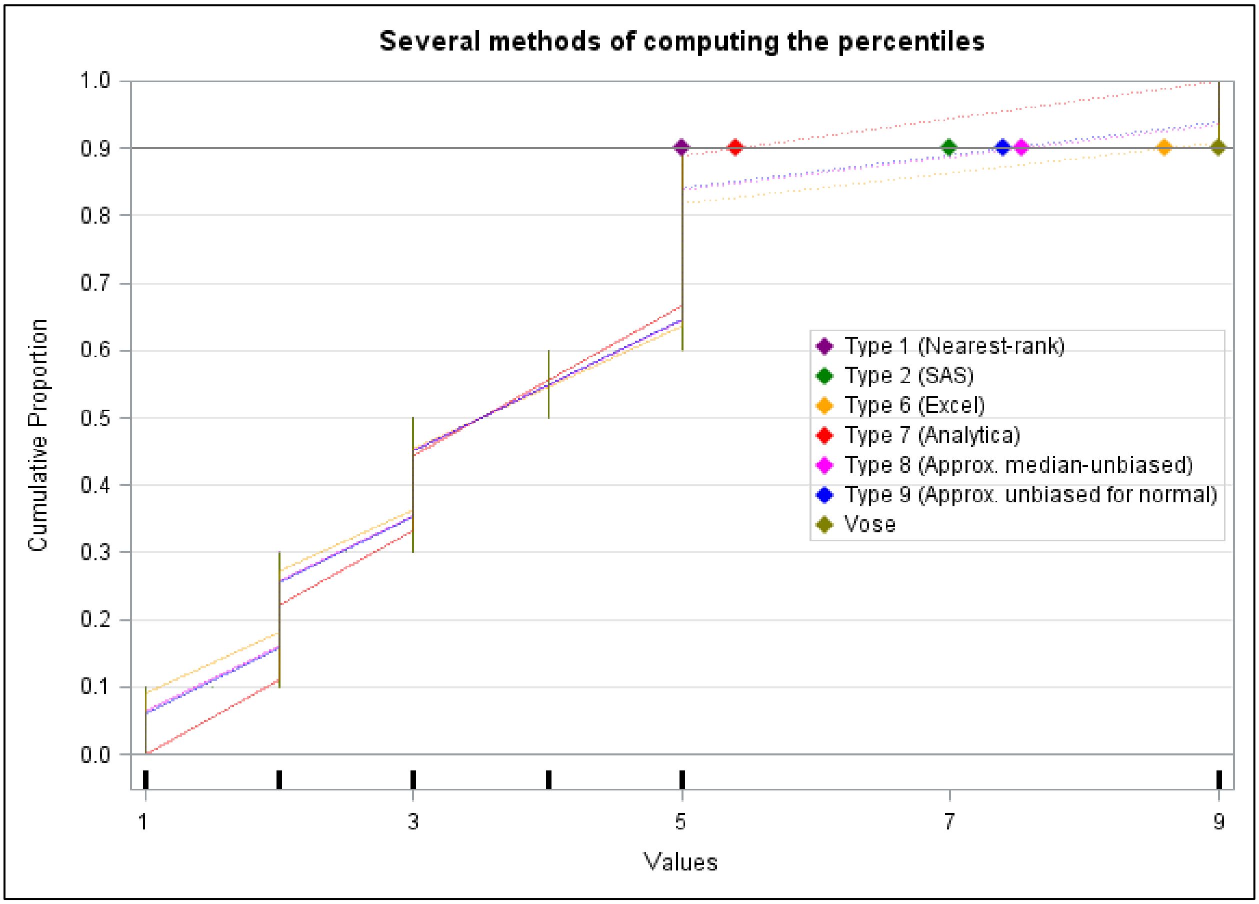

Figure 1. Several methods of computing percentiles for a small data set; 90th percentiles are shown.

To better understand the behavior of some of the methods used to compute percentiles in different software packages, two simulations were conducted for the most common distributions used in NIOSH-IREP. In both simulations, the percentiles obtained using the chosen methods are compared among themselves and against theoretical percentiles.

In the first simulation, five different continuous distributions were tested (these are the most common distributions used in the NIOSH-IREP software), each with different parameters settings, for a total of 10 combinations: Uniform U(1,6), Triangular T(1,3,5), Triangular T(1,5,5), Weibull W(2,1), Weibull W(1,10), Normal N(5,1), Normal N(25,5), Lognormal LN(1,3), Lognormal LN(3,6), and Lognormal LN(3,10).

There are different sampling methods that can be used to take random samples from a known distribution, and the most common methods are the Median Latin Hypercube (MLH), the Random Latin Hypercube (RLH), and the Monte Carlo (MC) sampling. In the first simulation, only the MLH and RLH sampling method were used.

To analyze the distribution of values for each of the percentile definitions used, multiple repetitions were used to take samples from each distribution. For the RLH sampling, $R = 500$ repetitions were used to assess a distribution of percentiles computed from samples. For MLH sampling, the resulting samples of the same size always have the same values when sorted. As a result, the percentiles are the same for all MLH samples of the same size, and thus only one repetition was performed for MLH sampling.

Six different sample sizes were used: 2,000, 5,000, 10,000, 20,000, 30,000, and 50,000 iterations. A sample size of 2,000 is currently used as default in NIOSH-IREP, and a sample size of 10,000 is used in NIOSH-IREP Enterprise Edition, where 30 runs are performed.

Six different methods of computing percentiles were chosen for this simulation, including the methods implemented in the most common software packages for Monte Carlo simulations: Type 1 method (Nearest-Rank method), Type 2 (default method in SAS software), Type 4 (used in @Risk software), Type 6 (used by Excel PERCENTILE.EXC function[1]), Type 7 (used in Analytica software), and also the method used by the Vose software.

Figure 2 shows box plots for the 99th percentile computed based on the six different methods, using MLH and RLH sampling, at different sample sizes, sampled from the LN(3,6) distribution. The method used by the SAS software has the smallest bias in estimating the 99th percentile at each sample size. Three of the methods used (Nearest-Rank, @Risk, and Analytica) underestimate the 99th percentile computed from the samples when compared with the theoretical 99th percentile of the distribution, and two methods (Excel, and Vose) overestimate the 99th percentile.

[1] In Excel, the PERCENTILE function is also commonly used, and that percentile calculation method is the same with that in Analytica (Type 7). For the purposes of the simulations in this paper, results using the newer percentile function PERCENTILE.EXC (Type 6) are reported as Excel estimates. Results from Excel using the PERCENTILE function are the same as those using Analytica.

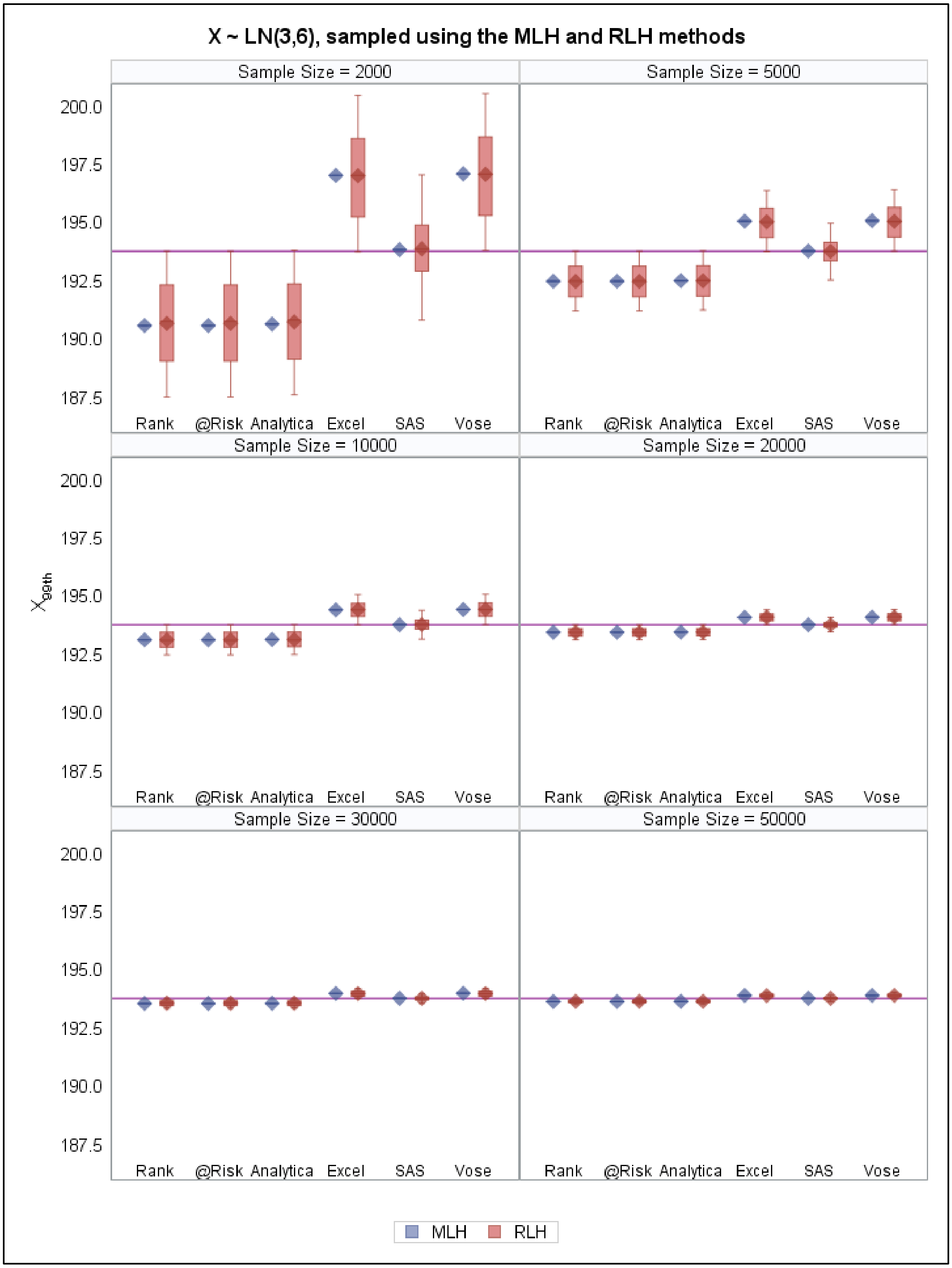

Figure 2. The 99th percentile computed using 6 different methods, for the LN(3,6) distribution.

Table 2 show the absolute bias at three different sample sizes (2,000, 10,000, and 20,000 repetitions), based on the method used in the Analytica software of estimating the 99th percentiles (Type 7), for the 10 simulated scenarios. Results in Table 2 are obtained using the MLH method, but the RLH method produces almost identical results. For all the distributions tested and at all sample sizes, the method used by Analytica to estimate the percentiles produces negative biases, meaning that this method is underestimating the theoretical population quantiles. As expected, the skewed distributions with the largest uncertainty have the largest bias, and the bias decreases as the sample size increases.

Table 2. Bias at different sample sizes, based on MLH sampling and the method used in the Analytica software for the estimation of the 99th percentiles.

Distribution	Bias at 2,000 sample size	Bias at 10,000 sample size	Bias at 20,000 sample size
Uniform (1,6)	-0.0012	-0.0002	-0.0001
Triangular (1,3,5)	-0.0034	-0.0007	-0.0003
Triangular (1,1,5)	-0.0049	-0.0010	-0.0005
Weibull (2,1)	-0.0056	-0.0011	-0.0006
Weibull (1,10)	-0.2419	-0.0489	-0.0245
Normal (5,1)	-0.0091	-0.0018	-0.0009
Normal (25,5)	-0.0455	-0.0092	-0.0046
LogNormal (1,3)	-0.1280	-0.0259	-0.0130
LogNormal (3,6)	-3.1305	-0.6359	-0.3186
LogNormal (3,10)	-13.1705	-2.6806	-1.3433

Figure 3 shows the relative bias of the 99th percentile computed based on 10 different methods (the 9 methods from Hyndman and Fan (1996), and the method implemented in Vose software), at different sample sizes, sampled from the LN(3,6) distribution, and using the RLH sampling. The relative bias of the 99th percentile computed using the MLH sampling (not shown) is almost identical with the relative bias from the RLH sampling.

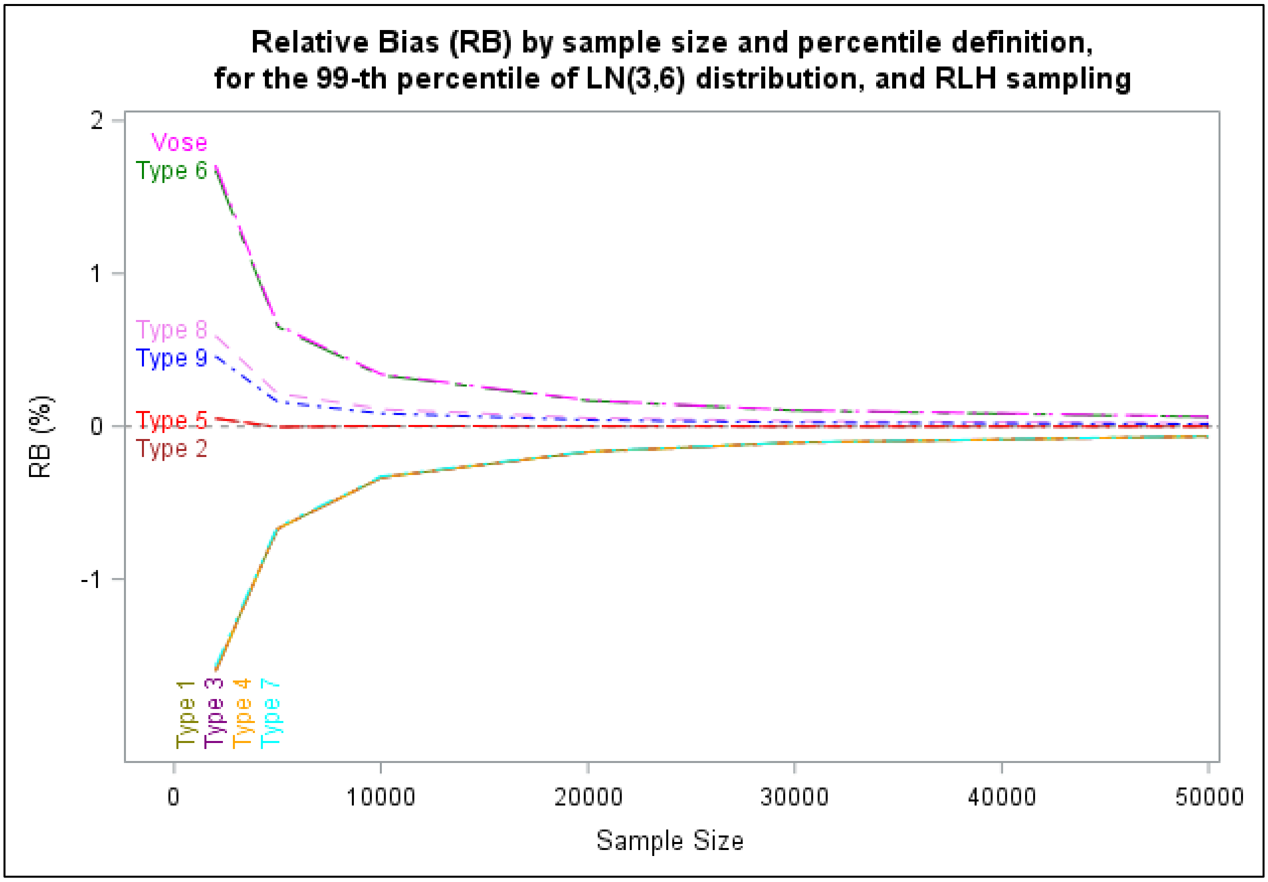

Figure 3. Relative Bias for the 99th percentile computed for 10 different percentile estimation methods.

From the tested methods used to compute the 99th percentile in the first simulation, the Type 2 and Type 5 methods seems to be the ones with the smallest relative bias and smallest relative root mean squared error at all the sample sizes, and for all the distributions scenarios tested, and the MLH sampling seems to have an advantage over the RLH sampling in terms of the RRMSE.

To investigate which method of computing the 99th percentile is better in terms of bias and variability when two distributions are combined, a second simulation was set up.

Since the normal and lognormal distributions are the most common distributions used in NIOSH-IREP, these distributions with either small or large standard deviations were used in this simulation, for both additive and multiplicative interactions, and a mixed interaction between the normal and lognormal distributions was also analyzed. A total of 20 scenarios were considered, obtained by either adding or multiplying two distributions from the following list: Normal (5,1), Normal (5,25), Lognormal (1,3), and Lognormal (3,6).

Figures 4 and 5 show the 99th percentile computed based on 2 different methods (Type 7 method used in Analytica, and Type 2 method, used by default in SAS), at different sample sizes, sampled from the Y1= X1 + X2 distribution, where X1 ~ LN(1,3), and X2 ~ LN(3,6), and for all the three sampling methods. The blue dots represent the average of the 99th percentile over the number of repetitions used, and the vertical red lines extend to one standard error around the average estimate. The red dotted horizontal lines in each plot represent the 95% confidence limits for the 99th percentile of the Y1= X1 + X2 distribution.

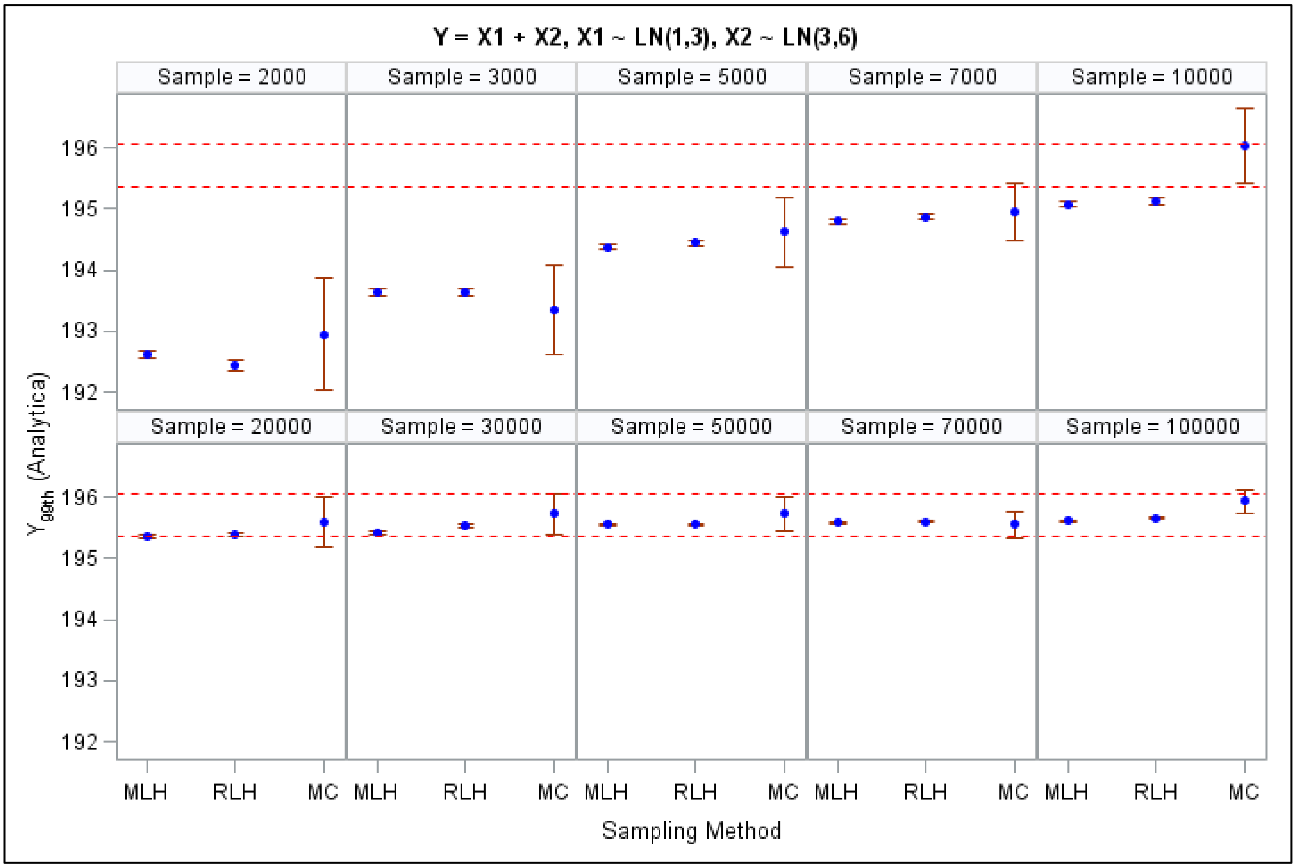

Figure 4. The 99th percentile computed using Type 7 method, for the sum of two distributions.

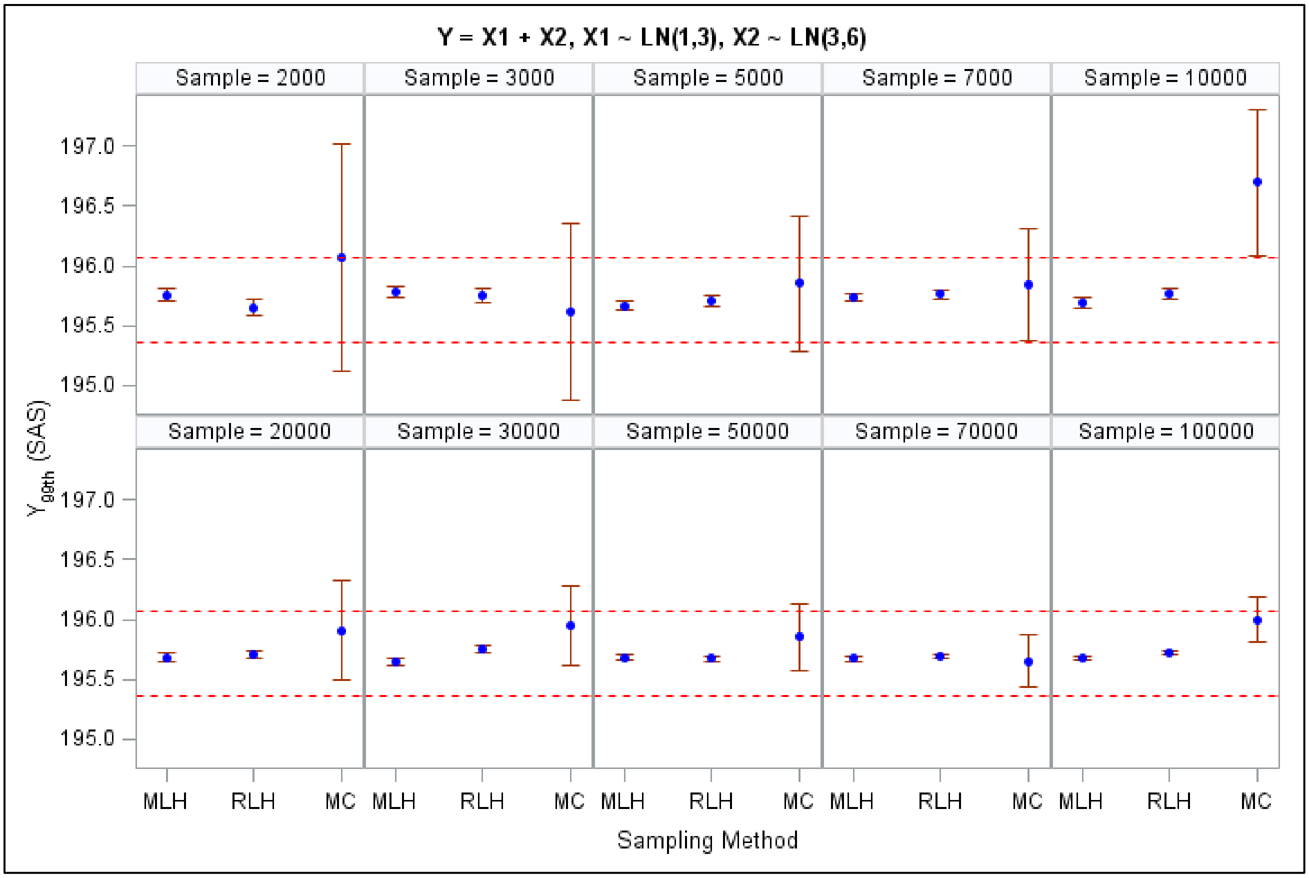

Figure 5. The 99th percentile computed using Type 2 method, for the sum of two distributions.

From the tested methods used to compute the 99th percentile in the second simulation, Type 2 method used by default in SAS, seems to be the one with the smallest relative bias and the MLH sampling seems to have an advantage over the RLH sampling in terms of the RRMSE for some of the scenarios tested. The Type 7 method used in Analytica underestimates the theoretical 99th percentile of the combined distribution in all scenarios, for most sample sizes, and for sample sizes up to 20,000 values the bias can be significant.

In conclusion, there is no standard definition of a percentile, however there are multiple definitions currently in use, with each software package using different methods to compute the percentiles. The performance of the percentile methods differs, depending on the percentile level, sample size, type of distribution, magnitude of uncertainty and of positive skewness; however, the larger the sample size, the less important the effect of the percentile definition used is. This section summarized and compared several percentile definitions and clarified which of these definitions is used in the NIOSH-IREP software, for the purpose of computing the PC values.

For all the simulations scenarios tested, the Type 2 method, used by default in the SAS software, seems to be the method with the smallest relative bias and smallest relative root mean squared error at all the sample sizes, in estimating the theoretical 99th percentile.

The Type 7 method, currently used in Analytica, underestimates the theoretical 99th percentile in all the scenarios tested, especially for sample sizes less than 20,000 values.

3.0 EFFECT OF AN ALTERNATIVE PERCENTILE ON PC VALUES

Based on the two simulations described before, it was determined that Type 2 percentile definition, used by default in the SAS software, is the method with the smallest relative bias and smallest relative root mean squared error at all the sample sizes, in estimating the theoretical 99th percentile. The Type 7 percentile definition, currently used in Analytica, and in the NIOSH-IREP software, will be denoted as the Default percentile, and the Type 2 percentile definition, used in the SAS software, will be denoted as the Alternative percentile. This section analyzes the effect of using the Alternative percentile instead of the Default percentile on the 99th PC values (99th percentile of PC).

Five sets of claims were selected from the NOCTS database to test the effect of using the Alternative percentile instead of the Default percentile (query date: December 17, 2019). Each set of cases were randomly selected based on the following ranges of the 99th PC values:

- Set 1: 45 single cancer cases with 99th PC values between 44.00% and 44.99%
- Set 2: 25 single cancer cases with 99th PC values between 45.00% and 48.99%
- Set 3: 22 single cancer cases with 99th PC values between 49.00% and 49.99%
- Set 4: 36 multiple cancer cases with Combined 99th PC values between 49.00% and 49.99%
- Set 5: 25 single cancer cases with 99th PC values between 50.00% and 51.99%.

Only partial results for Set 1, Set 3, and Set 4 are included in this report, with detailed results included in the main report.

Of the 283 single cancer cases with 99th PC values of record between 44.00% and 44.99%, 45 cases were randomly selected in Set 1 for testing purposes. Figure 6 shows the 99th PC values computed using both the Default percentile and the Alternative percentile for all 45 cases at 2,000 iterations. Since all cases were rerun on the newer IREP v.5.8.2, only a fraction of the recomputed PC values falls in the 44.00% and 44.99% range, with most of them being outside this range. Results obtained using the Alternative percentile are greater than those obtained using Default percentile for all cases at 2,000 iterations. The difference between the Alternative percentile and the Default percentile has an average increase of 0.27, with a maximum increase of 1.46 at 2,000 iterations.

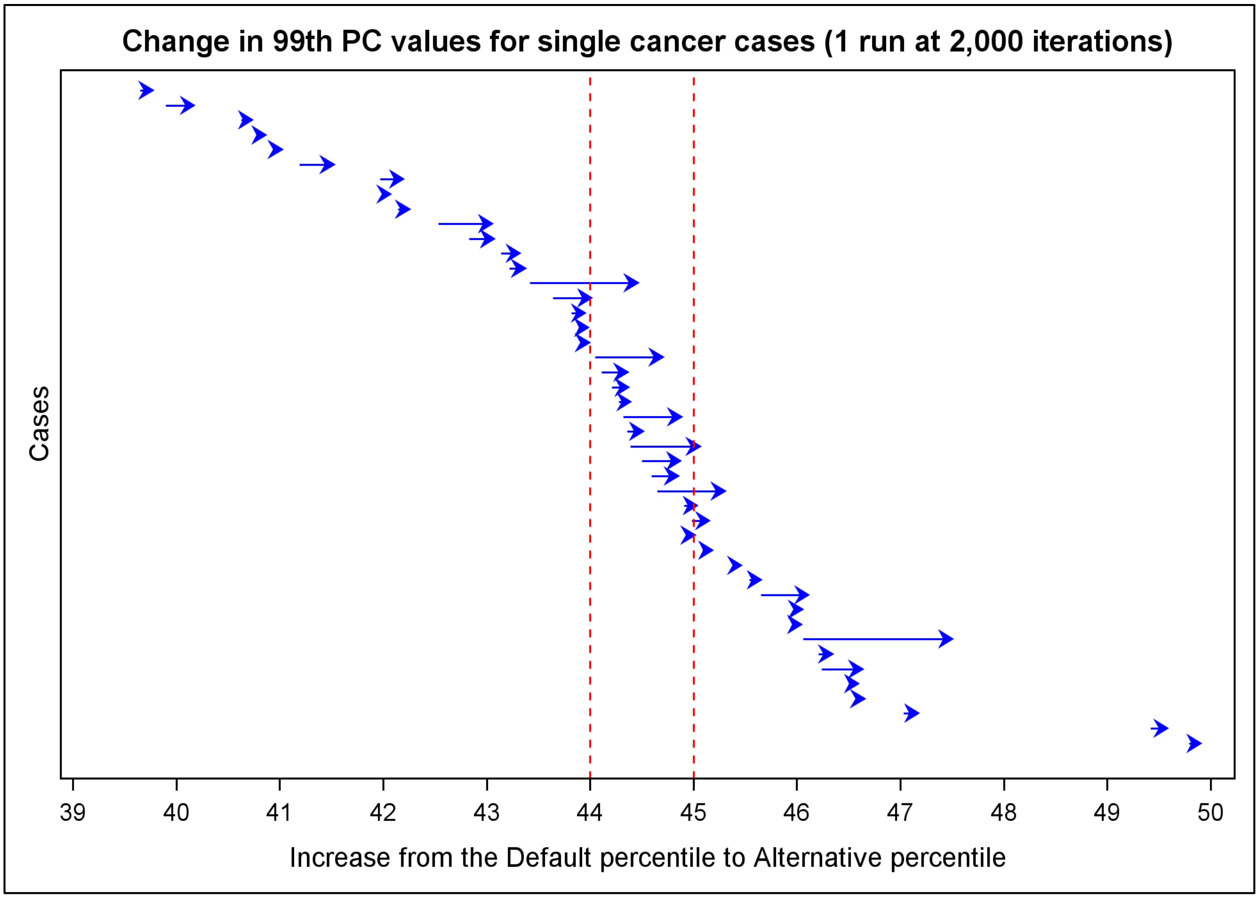

Figure 6. The increase in the 99th PC values from the Default percentile to the Alternative percentile at 2,000 iterations, for the 45 cases from Set 1 (PC values between 44.00% and 44.99%).

All 22 single cancer cases with 99th PC values between 49.00% and 49.99% were selected in Set 3 for testing purposes. The Alternative percentile produces greater 99th PC values than the Default percentile for all cases and each number of iterations. The difference between Alternative percentile and the Default percentile has an average increase of 0.04, with a minimum increase of 0.02, and maximum increase of 0.07 at 10,000 iterations, an average increase of 0.02, with a maximum increase of 0.04 at 20,000 iterations, and an average increase of 0.015, with a maximum increase of 0.02 at 30,000 iterations. Figure 7 shows the Avg. 99th PC values computed using both the Default percentile and the Alternative percentile for all 22 cases, based on 30 runs at 10,000 iterations.

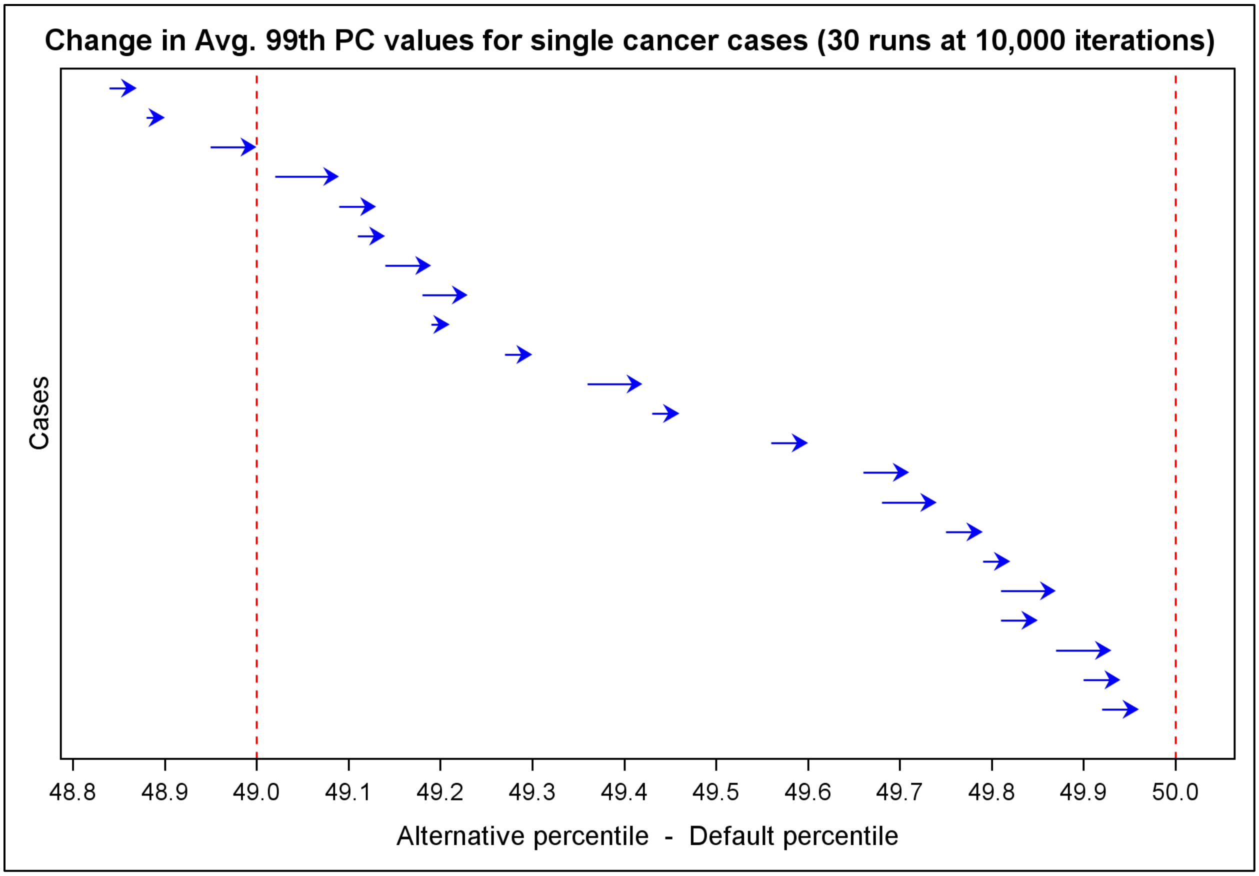

Figure 7. The increase in the Avg. 99th PC values from the Default percentile to the Alternative percentile, for the 22 single cancer cases from Set 3 (PC values between 49.00% and 49.99%), based on 30 runs at 10,000 iterations.

Figure 8 shows the Avg. 99th PC values computed using both the Default and the Alternative percentiles for all 22 cases based on 30 runs at 20,000 iterations. The lengths of the arrows in Figure 8 are smaller than in Figure 7, showing smaller increases in PC values at 20,000 iterations compared to 10,000 iterations. By running the 22 cases at 20,000 iterations, two of the 22 cases reach or surpass the 50.00% PC threshold using either the Default percentile or the Alternative percentile.

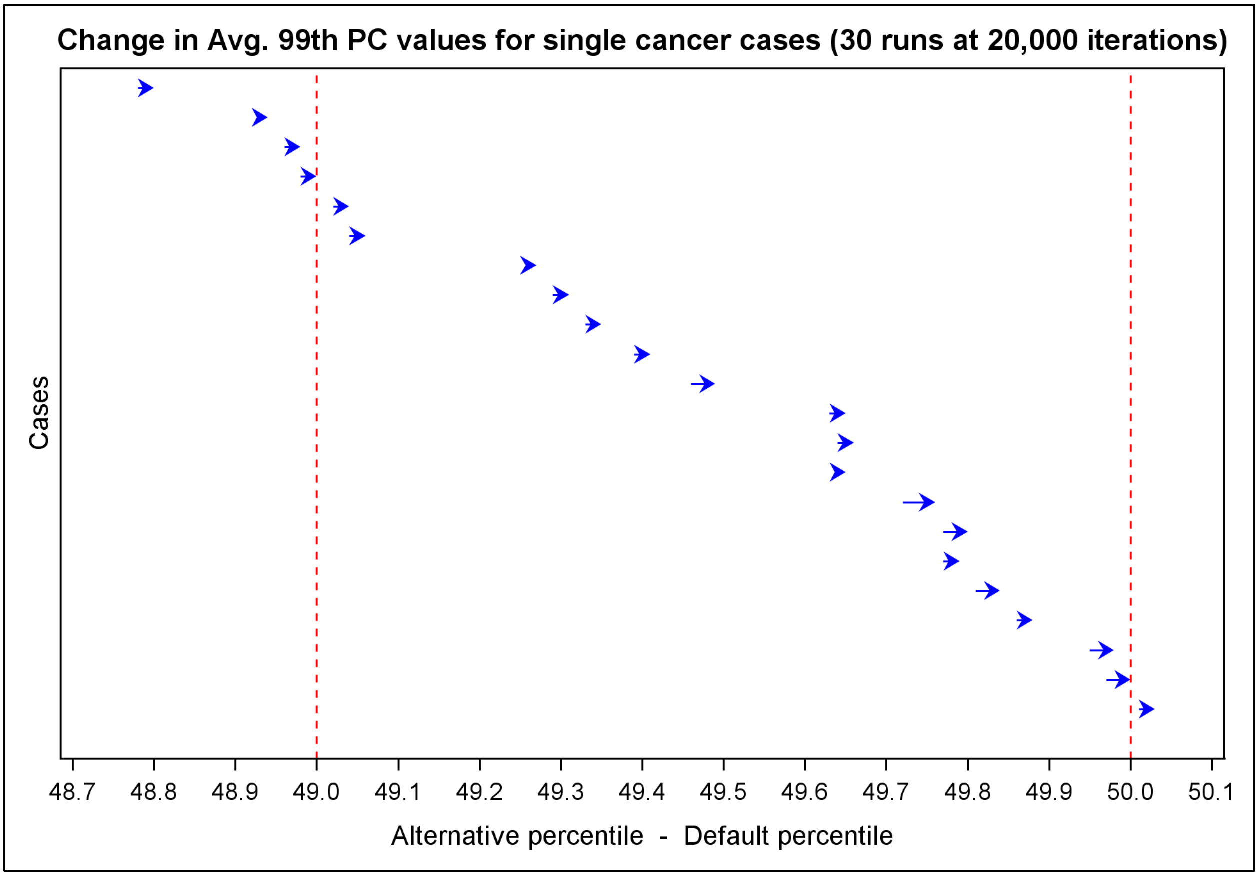

Figure 8. The increase in the Avg. 99th PC values from the Default percentile to the Alternative percentile, for the 22 single cancer cases from Set 3 (PC values between 49.00% and 49.99%), based on 30 runs at 20,000 iterations.

Figure 9 shows the Avg. 99th PC values computed using both the Default and the Alternative percentiles for all 22 cases based on 30 runs at 30,000 iterations. The lengths of the arrows in Figure 9 have slightly smaller magnitudes than the lengths of the arrows in Figure 8, showing smaller increases in PC values at 30,000 iterations compared to 20,000 iterations. By running the 22 cases at 30,000 iterations, two of the 22 cases reach or surpass the 50.00% PC threshold using either the Default percentile or the Alternative percentile.

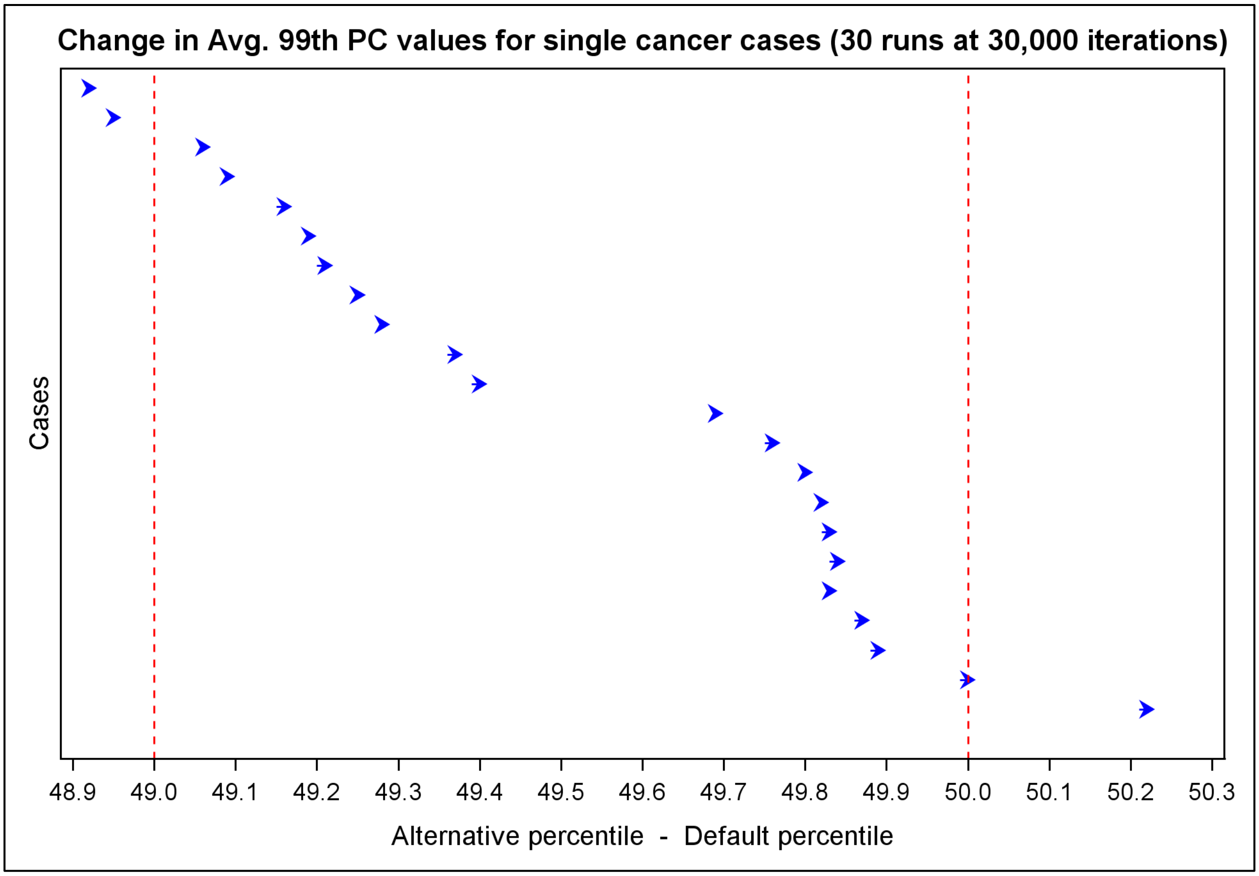

Figure 9. The increase in the Avg. 99th PC values from the Default percentile to the Alternative percentile, for the 22 single cancer cases from Set 3 (PC values between 49.00% and 49.99%), based on 30 runs at 30,000 iterations.

In addition to the single cancer cases with 99th PC values between 49.00% and 49.99%, all the 36 multiple cancer cases with Combined 99th PC values between 49.00% and 49.99% were also selected in Set 4 for testing purposes. The number of cancers varied between 2 and 13 cancers, for the cases selected. The Alternative percentile produced 99th PC values greater than the Default percentile for all cases and each number of iterations, and the difference between Alternative and the Default percentile has an average increase of 0.05, with a minimum increase of 0.03, and a maximum increase of 0.09, at 10,000 iterations.

Figure 10 shows the Combined Avg. 99th PC values computed using both the Default and the Alternative percentiles for all 36 cases, based on 30 runs at 10,000 iterations. The lengths of the arrows in Figure 10 are larger than the lengths of the arrows in Figure 7, showing the larger increases in PC values for the multiple cancer cases versus the single cancer cases. As it can be seen from the plot, two cases reach or surpass the 50.00% PC threshold using the Alternative percentile.

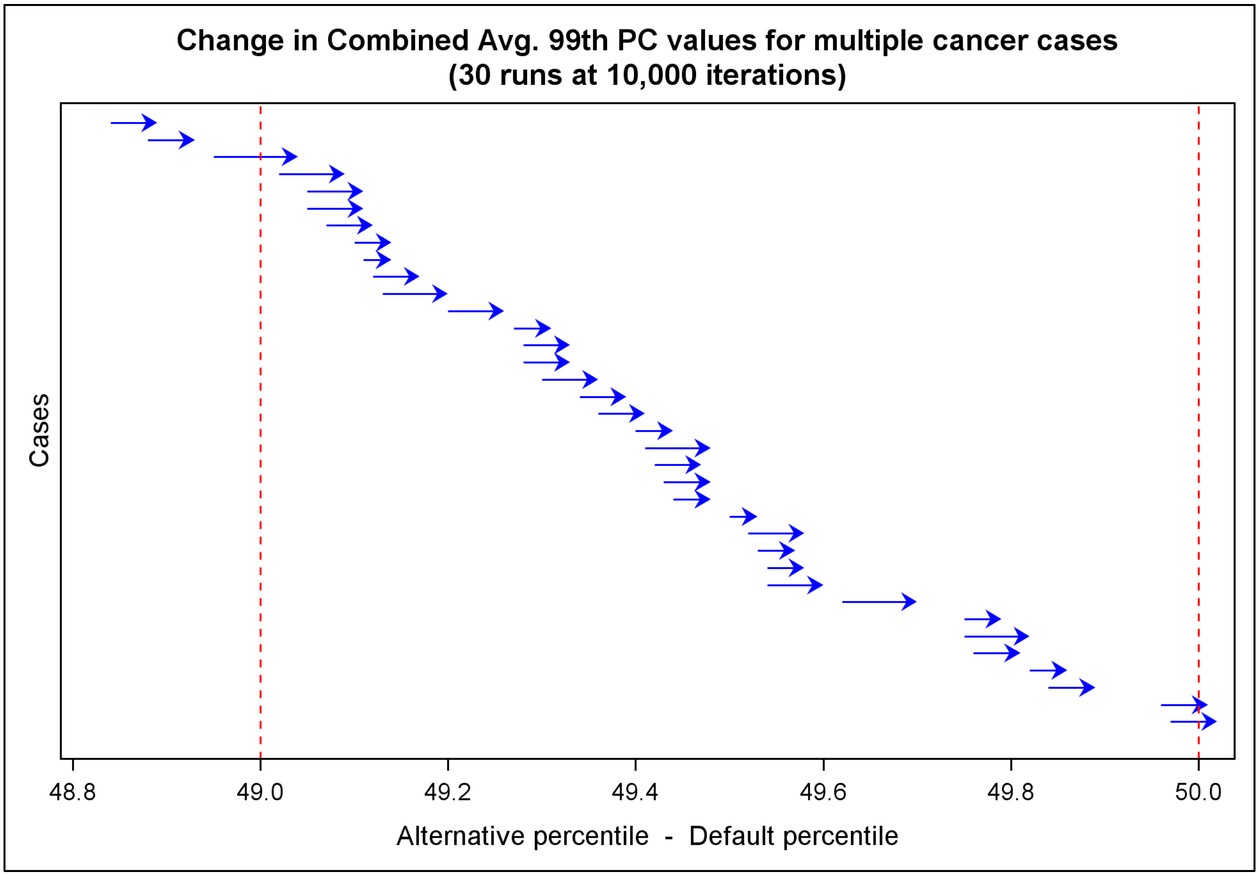

Figure 10. The increase in the Combined Avg. 99th PC values from using the Default percentile versus using the Alternative percentile, for the 36 multiple cancer cases from Set 4 (Combined Avg. PC values between 49.00% and 49.99%).

Figure 11 shows a plot with the running times at 10,000, 20,000, and 30,000 iterations versus the number of exposures for all 72 cases with Avg. 99th PC values between 45.00% and 51.99%. The largest number of exposures among the 72 cases is 724 exposures. The filled circles in the plot correspond to cases with Lung cancers, which usually take longer to complete because two lung cancer models are run in parallel. These running times were obtained using a research version of IREP EE, which is faster than the public IREP EE version; it is expected that the running times can be longer by up to one order of magnitude higher if cases are run on the current public IREP EE version.

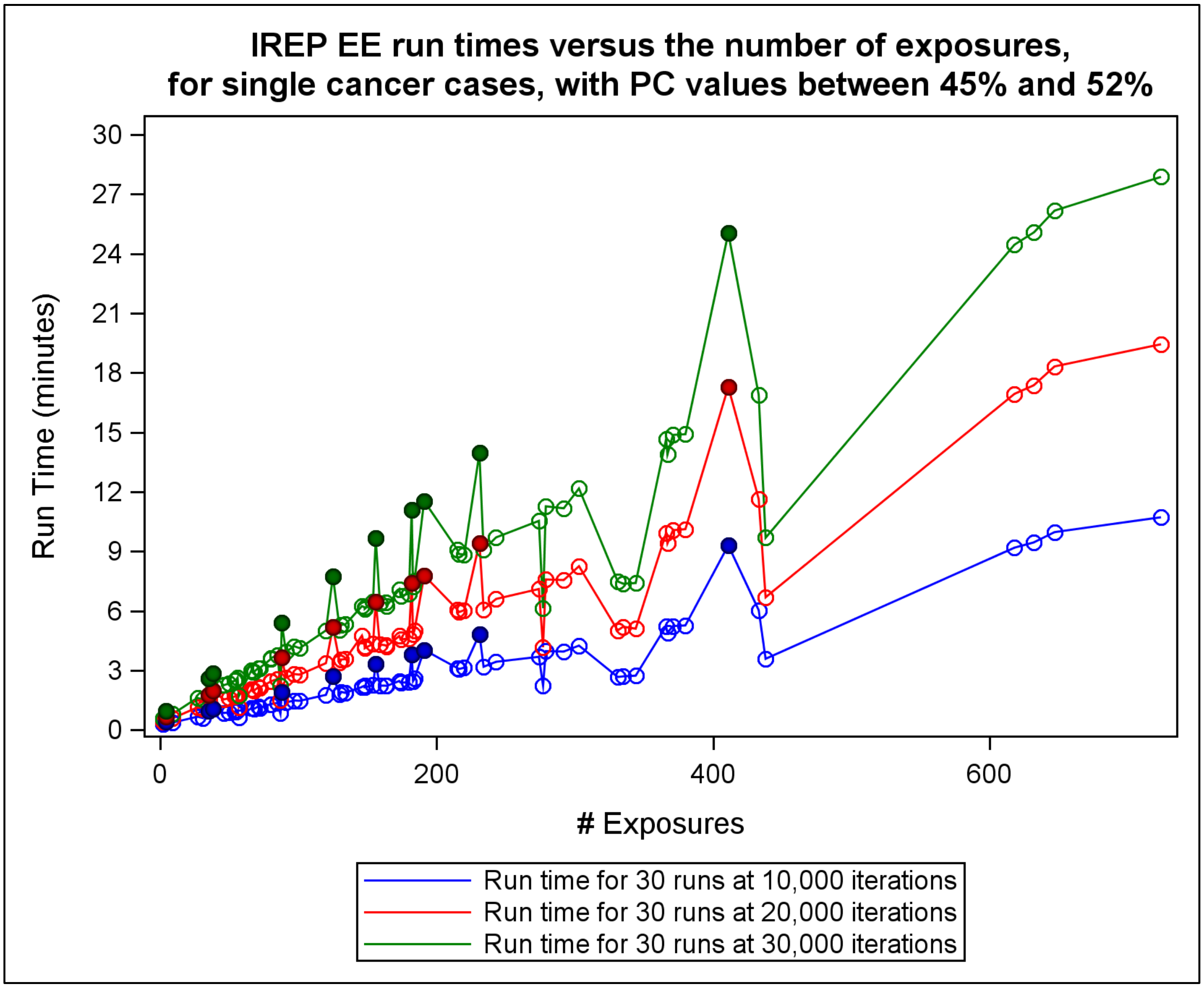

Figure 11. Run times (in minutes) at 10000, 20000, and 30000 iterations versus the number of exposures for the 72 cases with Avg. 99th PC values between 45.00% and 51.99%.

In conclusion, there are significant differences in the 99th PC values obtained using the Default percentile (computed using the Type 7 percentile definition, current default in the NIOSH-IREP software) and the Alternative percentile (computed using the Type 2 percentile definition) for the NOCTS cases tested, with the Alternative percentile always producing higher values than the Default percentile.

For the 45 cases with the 99th PC values between 44.00% and 44.99%, there is an average increase in the 99th PC values of 0.27, with a maximum increase of 1.46 at 2,000 iterations, when using the Default percentile versus using the Alternative percentile, an average increase of 0.06, with a maximum increase of 0.27 at 10,000 iterations, and an average increase of 0.02, with a maximum increase of 0.17 at 20,000 iterations.

For the 72 cases with the 99th PC values between 45.00% and 51.99%, there is an average increase in the 99th PC values of 0.04, with a maximum increase of 0.11 at 10,000 iterations, when switching from using the Default percentile to using the Alternative percentile. At 20,000 iterations there is an average increase of 0.02 (maximum increase of 0.05), while at 30,000 iterations the average increase is 0.015 (maximum 0.04).

To improve the estimates of 99th PC values, either the alternative percentile definition can be implemented in future NIOSH-IREP versions, or the number of iterations used should be increased, to minimize the underestimation effect on the 99th PC values.

The number of iterations used in the IREP version should be increased to a minimum of 10,000 iterations, but preferably to 20,000 iterations or more. The number of iterations used in the IREP Enterprise Edition should be respectively increased at a minimum of 20,000 iterations, but preferably to 30,000 iterations or more.

4.0 PC VARIABILITY AT DIFFERENT NUMBER OF ITERATIONS

Each case in NOCTS is run in IREP using 2,000 iterations and a default random seed of 99, and if the resulting 99th PC value is between 45.00% and 51.99%, the case was rerun on IREP EE with 30 runs at 10,000 iterations, and each run having a different random seed. Estimates of 99th percentile of PC obtained at 2,000 iterations can vary significantly depending on the random seed used. To better determine the variability of 99th PC values obtained at different numbers of iterations, an analysis was conducted for all cases with the 99th PC values between 45.00% and 51.99%. Altogether, a total of 77 cases with 99th PC values between 45.00% and 51.99% were selected for this analysis, and each of the 77 cases was run on IREP EE with 2,000, 10,000, 20,000, and 30,000 iterations.

Individual 99th PC values obtained from each set of 30 runs have a considerable higher variability at 2,000 iterations, and variability decreases as the number of iterations increases to 10,000, 20,000, and 30,000 iterations. For each set of 30 runs performed for a case at any number of iterations, the individual 99th PC values are approximately normally distributed, with the mean given by the Avg. 99th PC value.

Figure 12 shows the fitted normal distributions of individual 99th PC values at 2,000, 10,000, 20,000, and 30,000 iterations for one case with the Avg. 99th PC value at 30 runs and 10,000 iterations being 49.87%. The Avg. 99th PC values for each number of iterations are shown as vertical lines in the plot. This example shows that there is a high probability that one 99th PC value at 2,000 iterations could have been easily generated above 52%, making this case not eligible for a run on IREP EE (which would had been clearly an incorrect decision, and the compensability would had been incorrectly decided for this case).

Figure 13 shows the fitted normal distributions of the individual 99th PC values at 2,000, 10,000, 20,000, and 30,000 iterations for one case with the Avg. 99th PC value at 30 runs and 10,000 iterations being 50.47%. This example shows that there is a small probability that one 99th PC value at 2,000 iterations could have been generated below 45%, making this case not eligible for a run on IREP EE (which would had been an incorrect decision, since would had affected the compensability for this case).

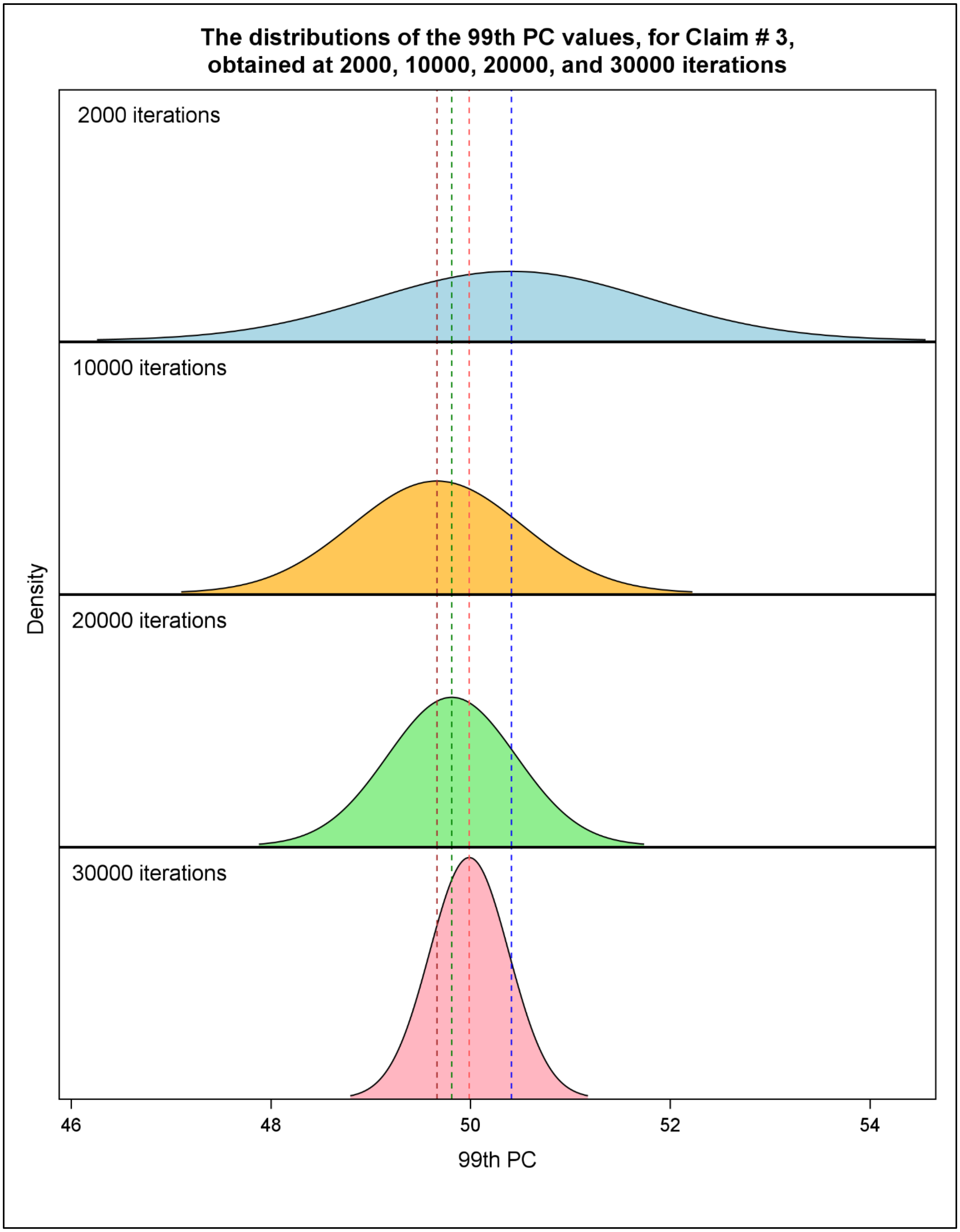

Figure 12. Fitted normal distributions for the 99th PC values at 2,000, 10,000, 20,000, and 30,000 iterations for a case with Avg. 99th PC value (30 runs at 10000 iterations) between 49% and 50%.

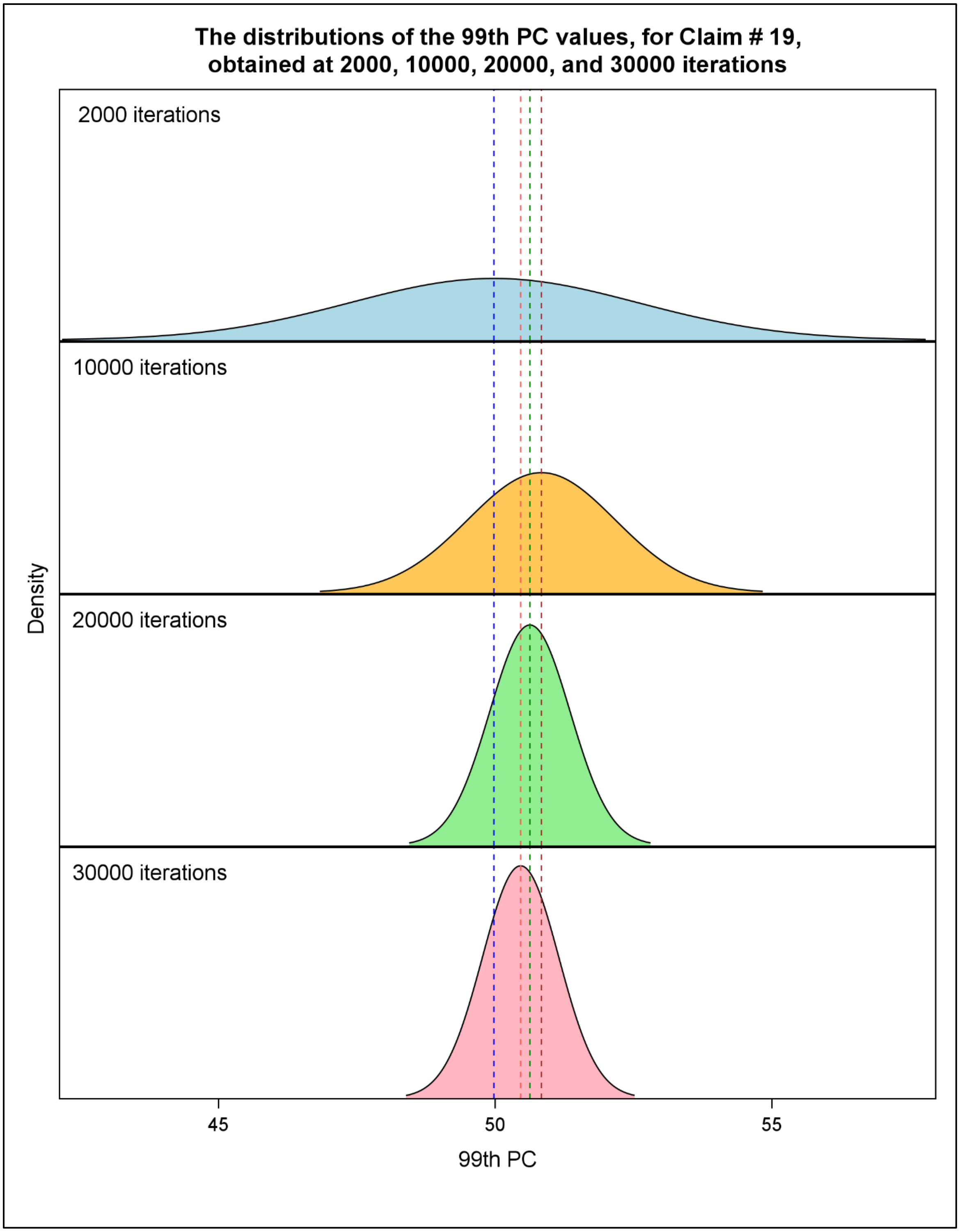

Figure 13. Fitted normal distributions for the 99th PC values at 2,000, 10,000, 20,000, and 30,000 iterations for a case with Avg. 99th PC value (10,000 iterations) between 50% and 52%.

The analysis of the 77 cases with PC values between 45.00% and 51.99% shows that there can be a large range of values for the individual 99th PC values, especially for the individual 99th PC values obtained at 2,000 iterations. While the individual 99th PC values have a very large variability, the Avg. 99th PC values have a much smaller variability, and so there is a big advantage in using the Avg. 99th PC values instead of the individual 99th PC values, due to the lower variability and the more stable values that are generated in the sampling process.

The examples shown in Figures 12 and 13 showed the peril of choosing one individual 99th PC value in determining a case outcome, and the fact that one individual 99th PC value obtained at the 2000 iterations might not always be a reliable indicator to choose which cases are run on the IREP Enterprise Edition, especially with the current settings of 2,000 iterations used in the regular IREP version. The current rule in use is that if the individual 99th PC value obtained at 2,000 iterations and default random seed is between 45.00% and 51.99%, corresponding to a range of (-5.0, 2.0) around the compensation threshold of 50.00%, then the case is rerun on IREP Enterprise Edition with 30 runs at 10,000 iterations. Based on the analysis with 76 cases (after excluding one outlier case), the range at 2,000 iterations around the 50% threshold to be used in selecting which cases need to be run in the IREP Enterprise Edition would need to be increased to (-8.8, 8.8). If the 99th PC value are obtained using the regular IREP version at 10,000 iterations, then the range used should be approximately (-4.4, 4.4), and if the 99th PC value are obtained using the regular IREP version at 20,000 iterations, then the range used should be approximately (-2.8, 2.8).

In conclusion, while the number of iterations used in the IREP and in the IREP Enterprise Edition versions need to be increased, the ranges around the 50% threshold to be used in selecting which cases need to be run in the IREP Enterprise Edition, based on 99th PC value obtained using the IREP version, should be adjusted accordingly, and several ranges are provided to be used depending on the number of iterations used in the IREP version.

5.0 PC CONVERGENCE AT 30 RUNS IN IREP ENTERPRISE EDITION

The convergence of the 99th PC values as well as of the Avg. 99th PC values, as the number of iterations increases, can be seen in Figure 14, which shows the 99th PC values obtained for a case with the 99th PC value 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 (10,000, 20,000, 30,000, 50,000, 100,000, 150,000, 200,000 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 99th PC value will be. For the case shown in Figure 14, the 95% confidence interval for the 99th PC value based on a 1,000,000 sample is (49.63, 50.02), and the 95% confidence interval for the theoretical 99th PC value based on a 5,000,000 sample is (49.76, 49.93). As the sample size increases, the 95% confidence interval for the 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 sample is 0.39, and the length of the 95% confidence interval for the theoretical 99th PC value obtained from the 5,000,000 sample is 0.17, and these lengths remain almost unchanged if different random seeds are used.

Figure 15 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 Figure 14. 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 increasing. 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 16 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 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 99^{th} 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, which was obtained by combining 3 samples with 2,000,000 individual PC values, since regular IREP cannot directly compute the 95% confidence limits for a case with large number of exposures and for a large sample size, as it will run out of memory. The 95% confidence interval for the theoretical 99^{th} PC value is (49.96, 50.05), which shows that for this case, even with a large sample size, the compensability of the case might be hard to decide. Since the length of the 95% confidence interval for the theoretical 99^{th} PC value is 0.09, this case suggests that there is a limit of the precision with which the theoretical 99^{th} PC value for a case can be bounded.

In conclusion, there is a clear trend of the individual 99^{th} PC values, as well as the Avg. 99^{th} PC values, to converge to a stable value that represents the theoretical 99^{th} PC value of the distribution of PC values for that case. The theoretical 99^{th} PC value of the distribution of PC values for a case cannot be determined exactly, but a 95% confidence interval for the theoretical 99^{th} 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 99^{th} 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.

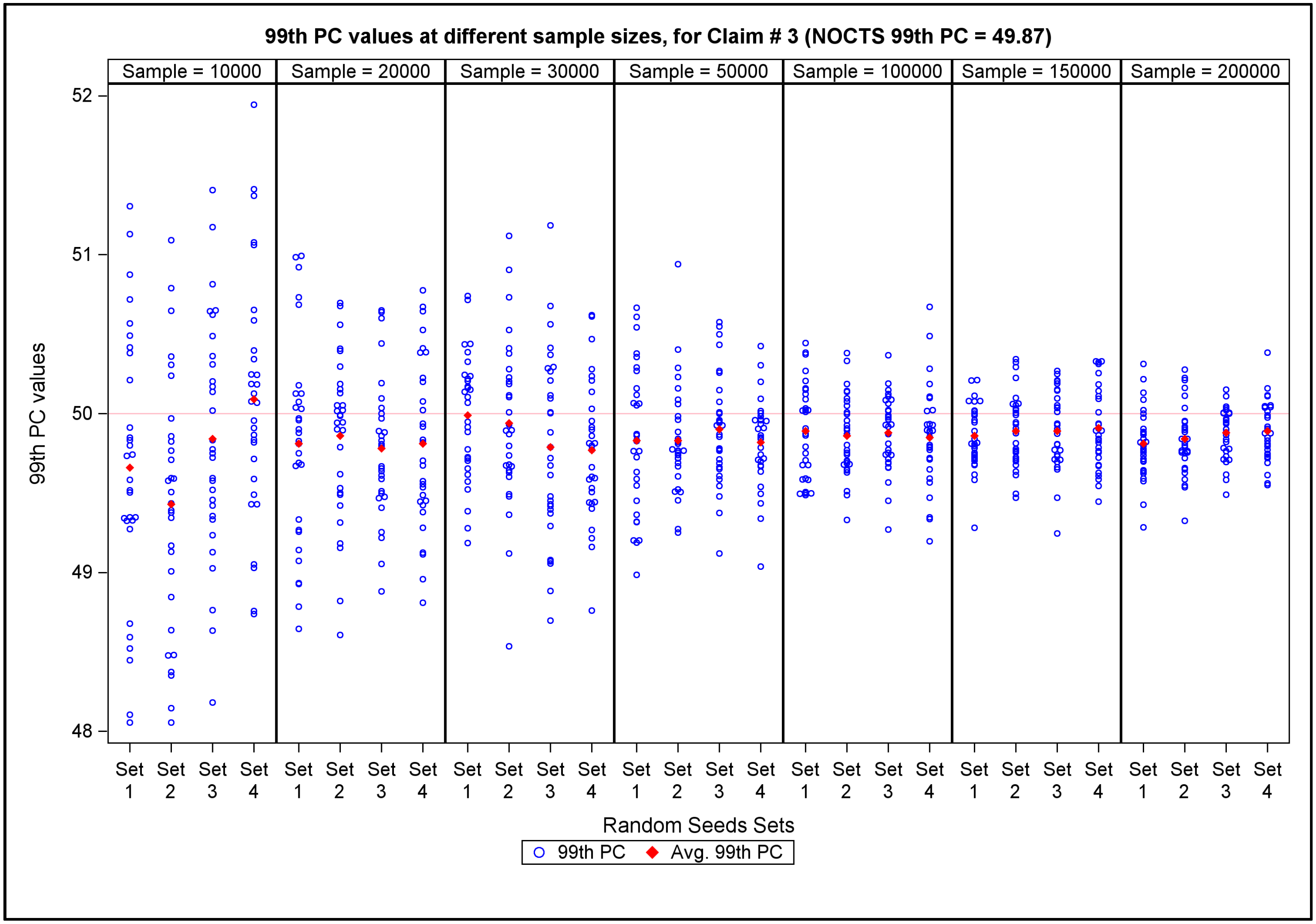

Figure 14. 99th PC values obtained at different sample sizes, for a case with the Avg. 99th PC value between 49% and 50%.

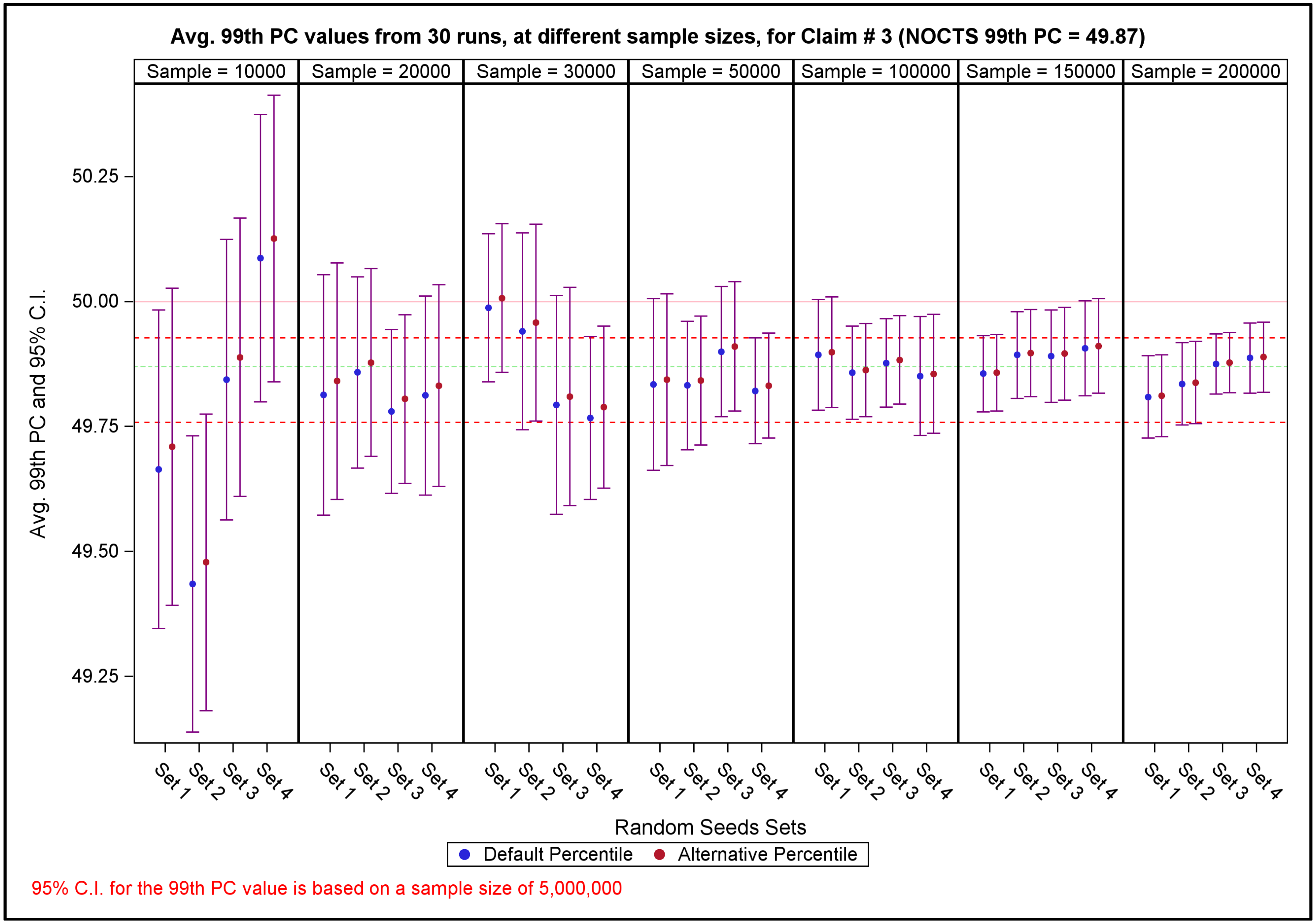

Figure 15. Avg. 99th PC values from 30 runs, and their 95% confidence intervals, obtained using the default and alternative percentiles.

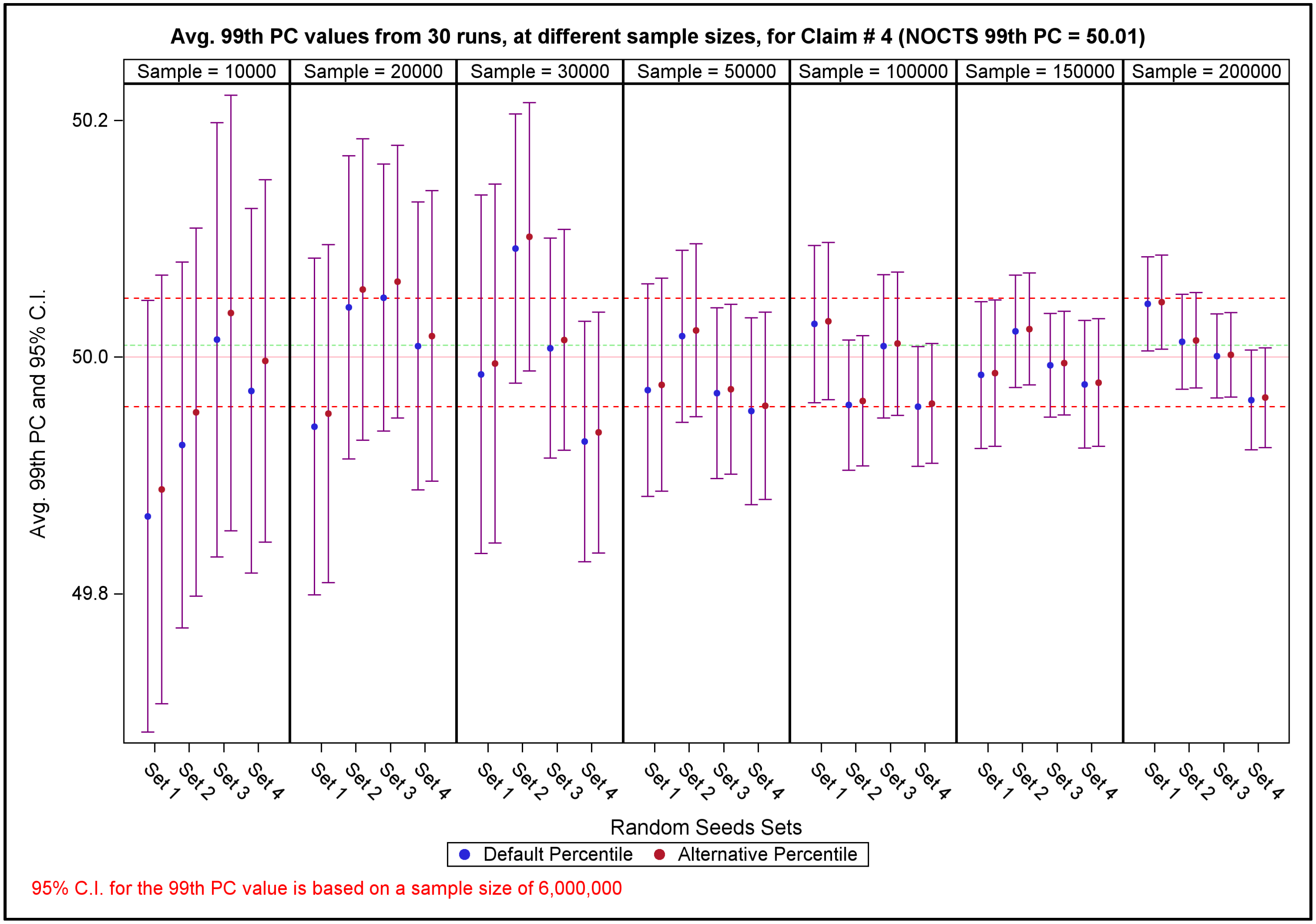

Figure 16. Avg. 99th PC values from 30 runs, and their 95% confidence intervals, obtained using the default and alternative percentiles.

6.0 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. For the Avg. 99th PC values to become relatively stable, 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 smaller or equal than a desired precision level (e.g., a length of the 95% confidence interval of 0.10 or 0.15), especially for the 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 previously 30 runs 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 to compute the Avg. 99th PC value for each combination of runs and iterations; ideally, a case should be completed in no more than 2 or 3 hours.

Twenty-three cases with the Avg. 99th PC values between 49.00% and 50.01% were selected for this analysis, and each case was 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 17 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 in 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 the analyses in which 30 runs were used in the IREP Enterprise Edition, even with 200,000 iterations per run, there were cases for which the lengths of the 95% confidence intervals were higher than 0.15, and none of 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.

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 18, which shows the Avg. 99th PC values obtained for the case shown in Figure 15. 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. 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 to converge 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 19 illustrates the convergence of the Avg. 99th PC values, as the number of iterations increases, for the case in Figure 16. 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 is less than 0.04. There is a clear trend of 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 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.

In conclusion, 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 number 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.

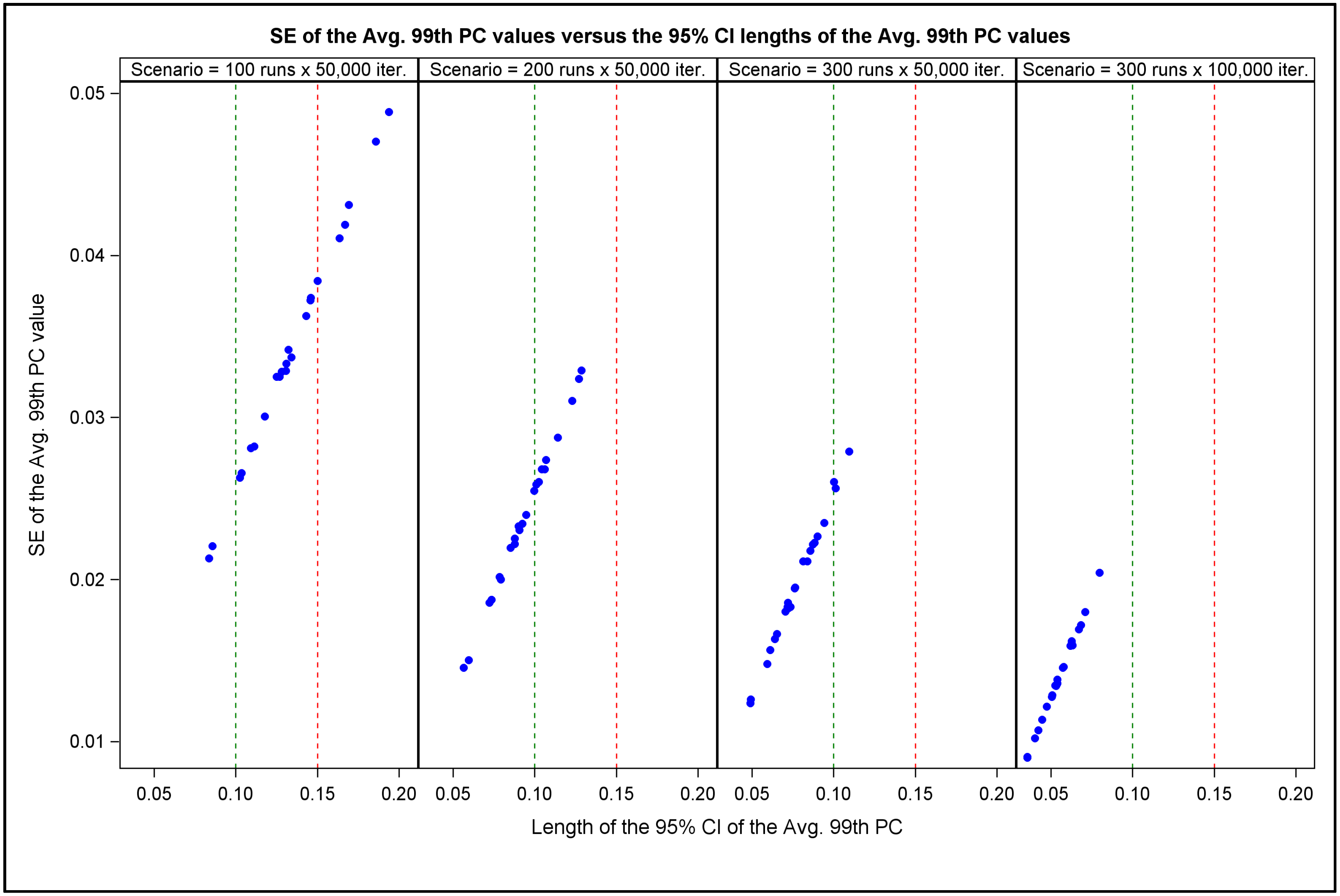

Figure 17. Standard errors (SE) versus the length of the 95% CIs from the four scenarios with different number of runs and iterations.

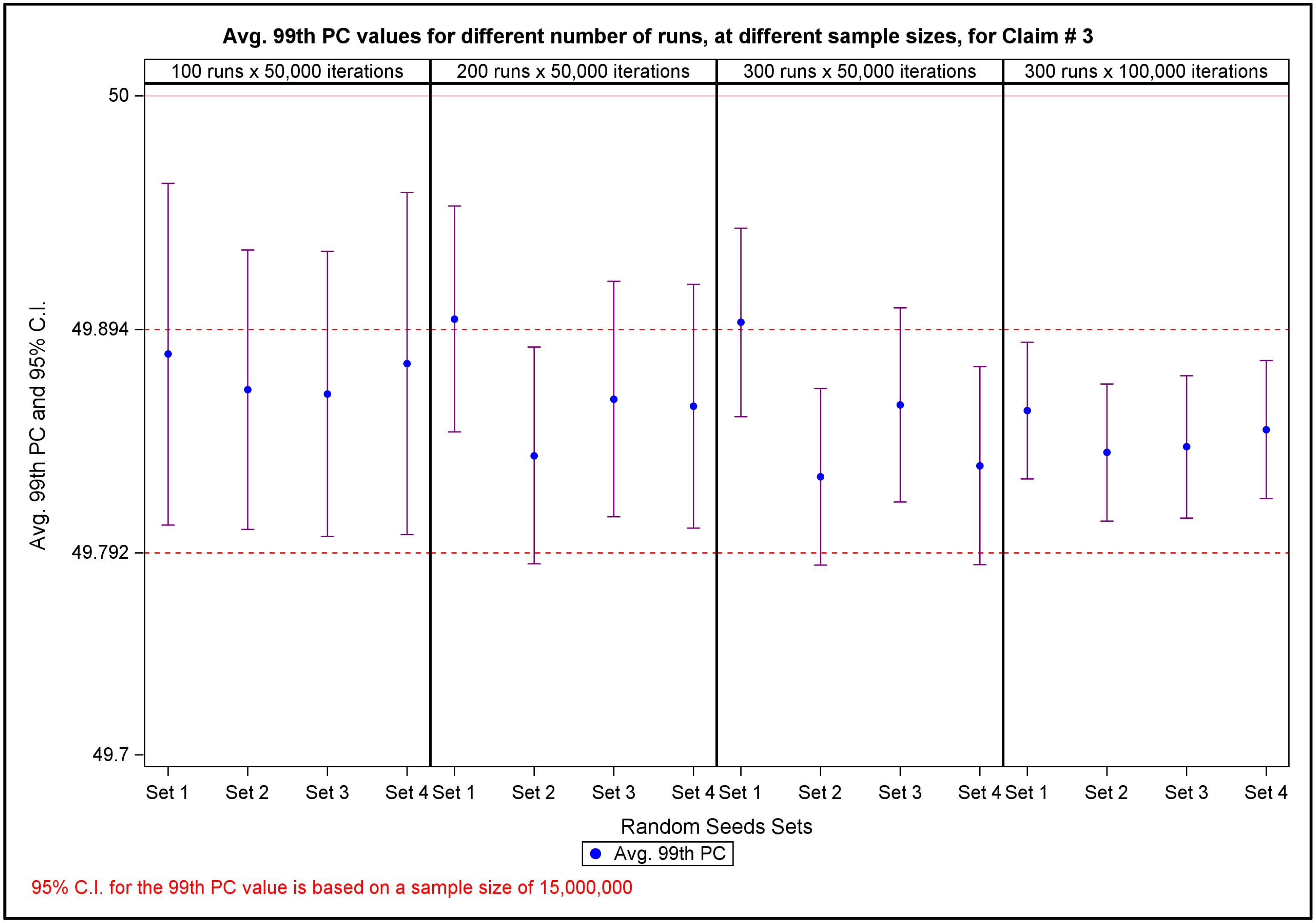

Figure 18. Avg. 99th PC values from 100, 200, and 300 runs, and their 95% CIs, and confidence limits for theoretical 99th PC value.

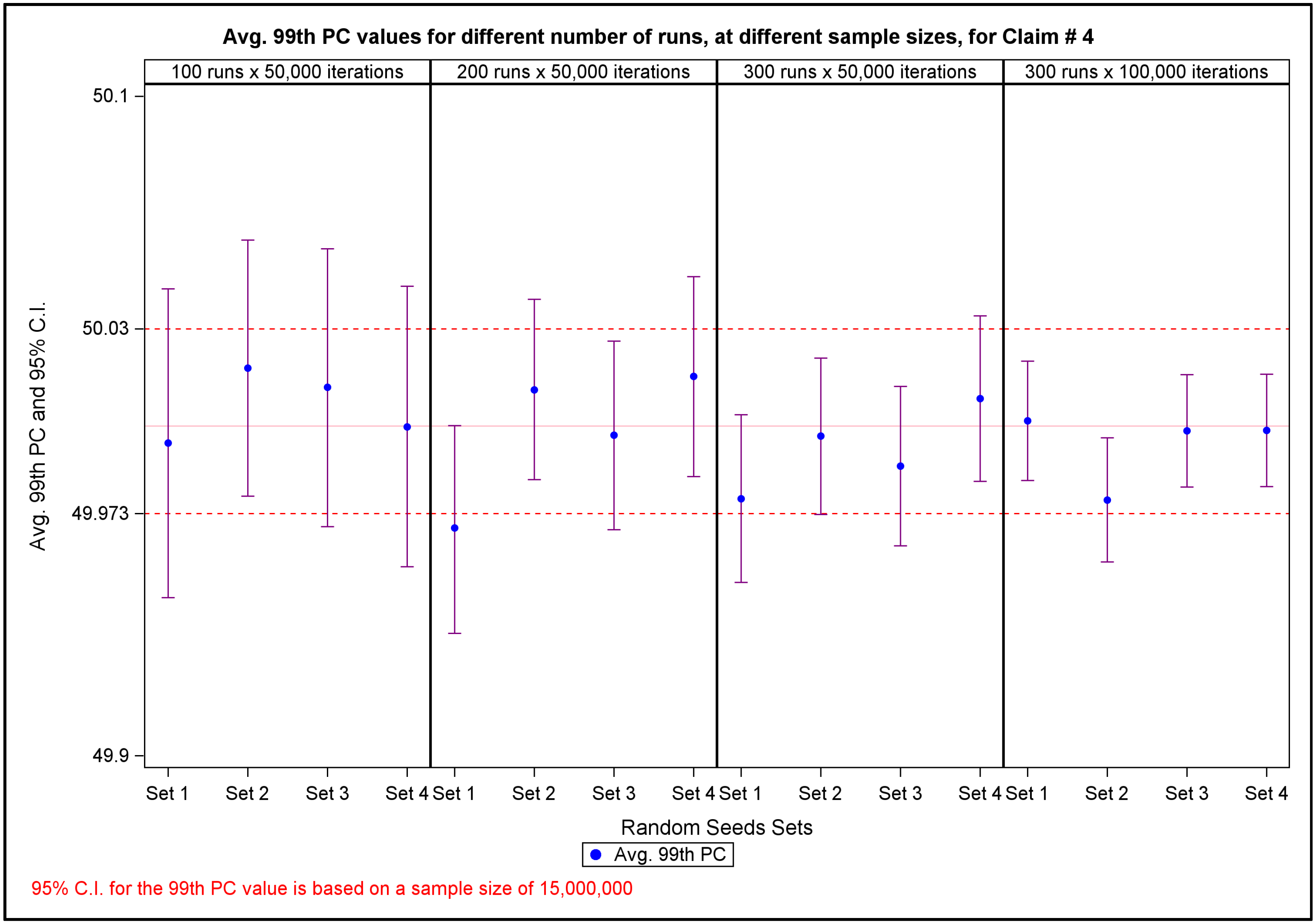

Figure 19. Avg. 99th PC values from 100, 200, and 300 runs, and their 95% CIs, and confidence limits for theoretical 99th PC value.

7.0 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, and 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 for 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 number of iterations for the tested cases. The function has an estimated exponent of -0.5. To use the IREP predictive tool, 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 precision level.
2) Run IREP Enterprise Edition using 300 runs and with 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, in order 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 (with 300 runs) times 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.

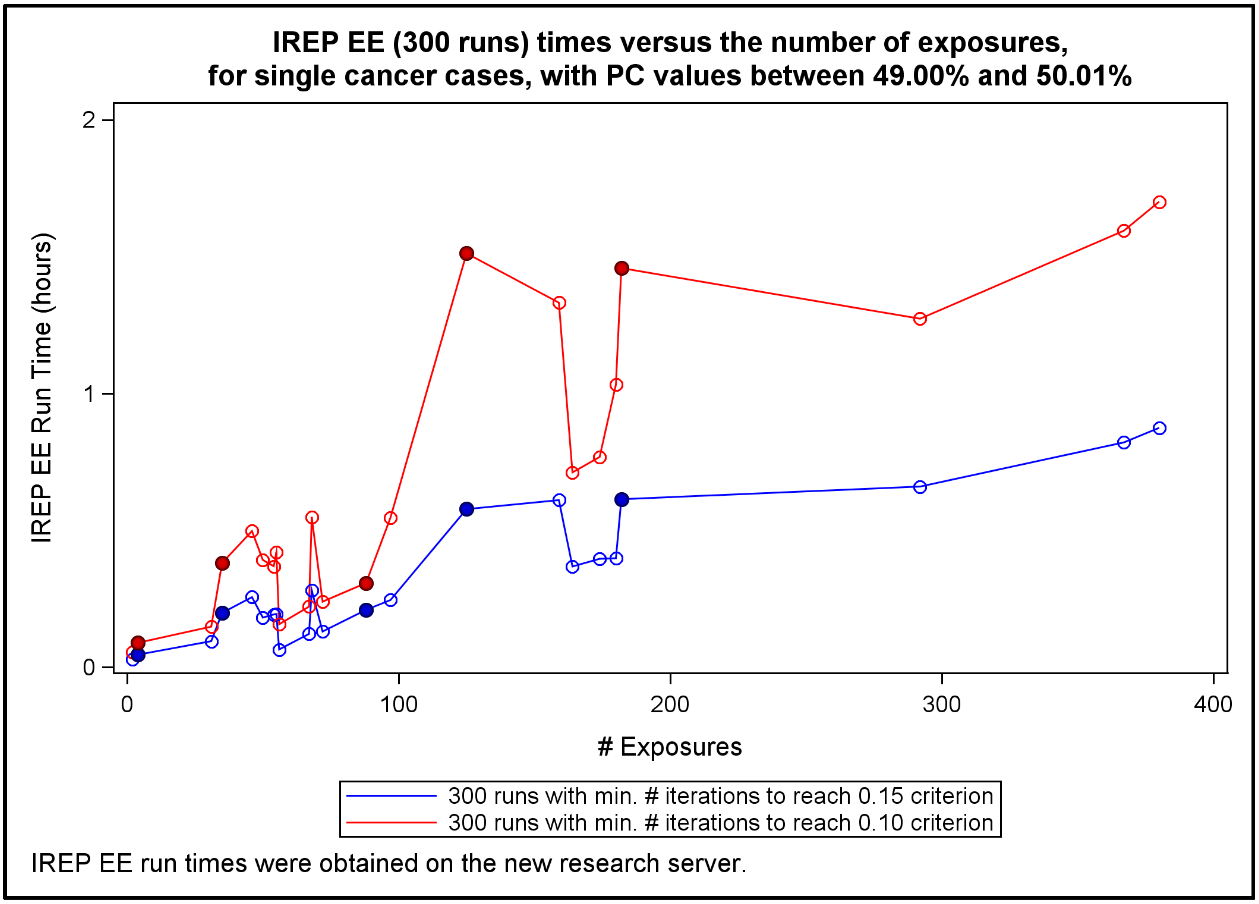

Figure 20. IREP EE running times for the 23 single cancer cases with Avg. 99th PC values between 49.00% and 50.01%.

Figure 21 shows the minimum number of iterations needed for each case to reach each of 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.

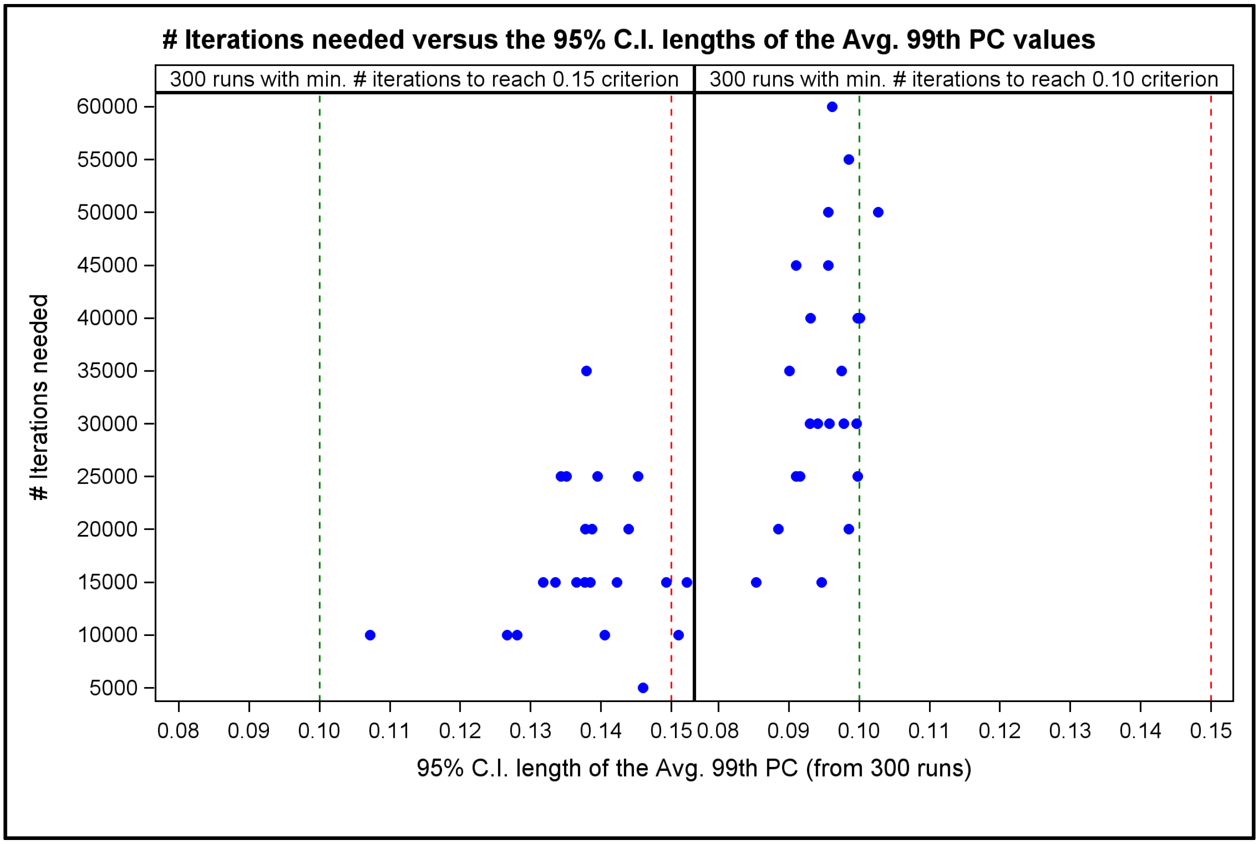

Figure 21. Minimum number of iterations necessary to reach the desired precision levels for 23 single cancer cases with 99th PC values between 49.00% and 50.01%.

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 the desired precision level for the 95% confidence intervals of 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 cases with Avg. 99^{th} PC values between 49.50% and 50.50%, one of the 5 scenarios of number of runs and iterations described in Table 3 can be used to reduce the variability of the Avg. 99^{th} PC values that is occurring by changing the sets of random seeds, and increase the accuracy of the Avg. 99^{th} PC value.

For cases with multiple cancers, a slightly modified procedure to use the IREP Predictive Tool will guarantee that the Combined Avg. 99^{th} PC value (obtained using the Multiple Cancer formula in IREP) will have the desired precision for the 95% confidence interval. This procedure relies on identifying the needed number of iterations for each cancer separately and using the largest of these number of iterations for all cancers.

In conclusion, an alternative way to reduce the length of the 95% confidence intervals for the Avg. 99^{th} PC values and also to minimize the IREP EE running times, is to use the IREP Predictive Tool to determine the minimum the number of iterations needed to achieve the desired precision in the Avg. 99^{th} PC values.

For the cases with PC values between 49.50% and 50.50%, one of the 5 scenarios of number of runs and iterations described in Table 3 is recommended to reduce the variability and increase the accuracy of the Avg. 99^{th} PC values. 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. 99^{th} PC values should be smaller than one of the desirable precision levels of 0.15 or 0.10.

By running a case with 300 runs and 50,000 to 100,000 iterations per run, we expect that the length of the 95% confidence interval around the Avg. 99^{th} PC values will be 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. 99^{th} PC values obtained in IREP EE can be estimated somewhere between 0.05 and 0.10, if enough number of iterations are used. This indicates there is a practical limit with which the precision around the Avg. 99^{th} 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 of cases, since it will require an extensive running time in IREP that becomes prohibitive.

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

As a validation exercise, comparisons between the 99th PC values obtained from the IREP version and the Avg. 99th PC values obtained from the IREP EE version were conducted for six different cases with small number of exposures and four different scenarios with a total of 300,000, 5,000,000, 10,000,000, and 15,000,000 iterations.

That analysis showed that as the number of total iterations increases, there are no significant difference between the 99th PC values obtained from a single run with very large number of iterations in IREP versus the Avg. 99th PC values obtained from multiple runs in IREP EE, for the same number of total iterations used.

Also, it was concluded that there is a slightly better precision in the 99th percentile of PC (smaller lengths for the 95% confidence interval in the Avg. 99th PC values obtained from IREP EE, versus the 99th percentile of PC from IREP) for the same number of total iterations used.

The IREP EE version has also the advantage that it can work with either 30 or 300 samples, and each of these samples can have 50,000 iterations or more, so overall 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.

9.0 RECOMMENDATIONS

Based on the analyses described in this report, the following actions are recommended regarding the settings that need to be changed in the current NIOSH-IREP versions, as well as related to how the cases are selected to be run on the two IREP versions.

1. The number of iterations used by default in the current IREP v.5.9 version should be increased from 2,000 iterations to 20,000 iterations.

2. The number of iterations used by default in the current IREP EE v.5.9 version should be increased from 10,000 iterations to 20,000 iterations.

3. The range for the cases that are selected to be run on the IREP EE version can be modified from the current range of (45.0% , 51.99%) to a range of (47.0% , 53.0%).

4. Incorporate the IREP Predictive Tool into the IREP EE version and change IREP EE to run cases with either 30 random seeds or 300 random seeds. The new IREP EE version would feel and operate just like the current IREP EE version, with no discernible difference for a user.

5. The IREP Predictive Tool implemented in IREP EE should be used for cases with Avg. 99^{th} PC values obtained from the IREP EE version in the range of (49.5%, 50.5%), to determine the minimum number of iterations required to achieve a precision of less than or equal to 0.10% around the Avg. 99^{th} PC value. Those cases will then be rerun on IREP EE with 300 runs and the number of iterations suggested by the IREP Predictive Tool.

REFERENCES

Hyndman R. J., Fan T., Sample Quantiles in Statistical Packages, The American Statistician, Vol. 50, No. 4, pp. 361-365, 1996.

Lumina Decision Systems Inc., User Guide Analytica 4.6, 13 May 2015, www.lumina.com.

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

Oracle, Crystal Ball 11.1.2 User Guide, 2010, www.oracle.com.

Palisade, @Risk 7.0 User Guide, July 2016, www.palisade.com.

Stancescu D., Apostoaei A. I., Thomas B.A., Comparison of Several Percentile Definitions, DCAS, December 2019.

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

Stancescu D., Apostoaei A. I., Thomas B.A., Increasing the Accuracy of the 99th Percentile of PC, DCAS, July 2022.

Vose Software, Model Risk 5.0 Help, 2012, www.vosesoftware.com.

LIST OF TERMS, ABBREVIATIONS AND ACRONYMS

Sampling Methods

MLH	Median Latin Hypercube sampling method
RLH	Random Latin Hypercube sampling method
MC	Monte-Carlo sampling method

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
T(min,mode,max)	Triangular distribution defined by parameters min, mode and max
min	Minimum of a Triangular probability distribution
mode	Mode of a Triangular probability distribution
max	Maximum of a Triangular probability distribution
U(min,max)	Uniform distribution defined by parameters min and max
min	Minimum of a Uniform probability distribution
max	Maximum of a Uniform probability distribution
W(shape,scale)	Weibull distribution defined by parameters shape and scale
shape	shape parameter of a Weibull probability distribution
scale	scale parameter of a Weibull probability distribution

Other terms

99^{th} PC value	99^{th} percentile of probability of causation (PC)
Avg. 99^{th} PC value	Average of the 99^{th} 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
DOL	Department of Labor
NIOSH	National Institute of Occupational Safety and Health
NOCTS	NIOSH DCAS Claims Tracking System
PC	Probability of Causation
RB	Relative Bias
RRMSE	Relative Root-Mean Square Error