Appendix K

Benchmark Small Sample Adjustment Tables

Benchmark Small Sample Adjustment Tables:
Derivation and Application Methods

Introduction

This appendix describes the method and rationale for the construction and application of benchmark small-sample adjustment tables. These tables provide allowances for performance results that fail benchmarks because of problems posed by small samples.

For performance to pass a strictly applied benchmark, results for small samples can often require perfect performance even though the benchmark allows less than perfect performance, such as 90 percent of OSS tasks completed on time. Additionally, with infrequent exception, strict application requires performance actually higher than the benchmark. This poses somewhat of a dilemma because, similar to the "alpha vs. beta" problem in statistical testing, solutions for one situation cause problems for the other situation. The parties have somewhat irreconcilable interests. The ILECs are concerned with erroneous decisions under conditions when an operational process actually allows "a meaningful opportunity to compete" as defined by the benchmark. The CLECs are concerned with erroneous decisions under conditions when an operational process actually does not allow "a meaningful opportunity to compete" as defined by the benchmark. The following examples illustrate two types of problems.

Example 1: Problems under "meaningful opportunity" (passing) conditions.

These problems exist when aggregate performance for all CLECs *allows* a meaningful opportunity to compete as illustrated by the following results for a 90 percent benchmark. Twenty CLECs submit five orders each for a total of 100 orders. We assume 10 "delinquent orders," that is, orders with necessary OSS tasks not completed within the benchmark time interval criterion. The process is in parity because 90 percent of the orders are completed within the time criterion. However, between 2 and 10 CLECs will *fail* the benchmark. The performance for at least two CLECs will fail as illustrated by the following outcome:

a. Two CLECs have no orders completed within the benchmark timeframe (zero percent performance result), accounting for the 10 delinquent orders (2 CLECs x 5 orders each).

The performance for as many as ten CLECs could fail the benchmark as illustrated by the following outcome:

b. Ten CLECs each have one order not completed within the benchmark timeframe (80 percent performance result), accounting for the 10 delinquent orders (10 CLECs x 1 order each).

These situations illustrate the logical underpinning of the small sample adjustment tables. In this illustration, a small sample table would allow one additional miss per CLEC result. This would allow the more likely outcomes (one delinquent order per CLEC) to be deemed in parity consistent with the 90 percent aggregate process performance level.

Example 2: Problems under “no meaningful opportunity” (failure) conditions:

These problems exist when aggregate performance for all CLECs *does not allow* a meaningful opportunity to compete, as illustrated with the following results, again for a 90 percent benchmark. Twenty CLECs submit five orders each for a total of 100 orders. We assume 20 delinquent orders. The process is failing because only 80 percent of orders are within the time criterion. However, between 10 and 20 (all) CLECs will *pass* the benchmark when a small sample table is used allowing one delinquent order per CLEC. The performance for at least ten CLECs will pass, as illustrated by the following outcome:

a. Ten CLECs have two orders not completed within the benchmark timeframe (sixty percent performance result), accounting for the 20 delinquent orders (10 CLECs x 2 delinquent orders each). The remaining ten CLECs would have no delinquent orders and thus would pass the benchmark.

The performance for as many as all twenty CLECs could pass the benchmark as illustrated by the following outcome:

b. Twenty CLECs each have one order not completed within the benchmark timeframe (80 percent performance result), accounting for the 20 delinquent orders (20 CLECs x 1 order each). However, all CLECs pass the benchmark because the small sample adjustment tables allow one "miss" each.

When strictly applied, the benchmarks actually require performance higher than the nominally specified percentage level as demonstrated in Table 1.

Table 1

(1) Sample size	(2) "Misses" permitted: Absolute benchmark	(3) Effective percentage: Absolute benchmark	(4) "Misses" permitted: Adjusted* benchmark	(5) Effective percentage: Adjusted* benchmark	(6) Sample size percentage	(7) Weighted effective percentage: Absolute benchmark	(8) Weighted effective percentage: Adjusted* benchmark
1	0	100.0%	1	0.0%	15.9	15.9	0.0
2	0	100.0%	1	50.0%	8.7	8.7	4.4
3	0	100.0%	1	66.7%	5.9	5.9	3.9
4	0	100.0%	1	75.0%	8.5	8.5	6.3
5	0	100.0%	1	80.0%	3.1	3.1	2.5
6	0	100.0%	1	83.3%	5.9	5.9	4.9
7	0	100.0%	1	85.7%	4.4	4.4	3.7
8	0	100.0%	1	87.5%	3.8	3.8	3.4
9	0	100.0%	1	88.9%	3.6	3.6	3.2
10	1	90.0%	1	90.0%	2.6	2.3	2.1
11	1	90.9%	2	81.8%	2.8	2.6	2.1
12	1	91.7%	2	83.3%	2.3	2.1	1.8
13	1	92.3%	2	84.6%	2.1	1.9	1.6
14	1	92.9%	2	85.7%	3.6	3.3	2.9
15	1	93.3%	2	86.7%	2.8	2.6	2.3
16	1	93.8%	2	87.5%	2.3	2.2	1.9
17	1	94.1%	2	88.2%	2.8	2.7	2.3
18	1	94.4%	2	88.9%	3.1	2.9	2.6
19	1	94.7%	2	89.5%	1.8	1.7	1.5
20	2	90.0%	2	90.0%	2.3	2.1	1.9
21	2	90.5%	3	85.7%	1.5	1.4	1.2
22	2	90.9%	3	86.4%	1.8	1.6	1.4
23	2	91.3%	3	87.0%	1.0	0.9	0.8
24	2	91.7%	3	87.5%	1.8	1.6	1.4
25	2	92.0%	3	88.0%	1.8	1.7	1.5
26	2	92.3%	3	88.5%	0.5	0.5	0.4
27	2	92.6%	3	88.9%	0.3	0.2	0.2
28	2	92.9%	3	89.3%	0.8	0.7	0.6
29	2	93.1%	3	89.7%	1.3	1.2	1.1
30	3	90.0%	3	90.0%	1.0	0.9	0.8
	Average =	94.5%		81.8%	Sum = 100.0	96.9	64.7

* Adjusted by adding one (1) additional permitted miss to each sample size result unless the original effective percentage equals the 90% nominal percentage.

Table 1 shows the effective percentage for the 90 percent benchmark for sample sizes of one (1) to thirty (30) both for an absolute benchmark application and an application allowing one additional "miss." For example, for a sample size of 19, since two "misses" out of 19 orders equals 89.5 percent on-time performance, it fails the benchmark. Therefore with an absolute application only one miss is allowed (column 2). Thus a performance result of at least 18 on-time orders out of 19 is required to pass – effectively a 94.7 percent performance requirement (column 3). On average, for sample sizes of one to thirty, a 90 percent benchmark requires 94.5 percent on-time performance to pass the benchmark. On the other hand, if one additional miss (column 4) was allowed so that the effective percentage was never greater than the nominal benchmark percentage, then this "adjusted" benchmark would only require 81.8 percent on-time performance for this range of samples (column 5). When these averages are adjusted for the fact that some sample sizes are more numerous than others, the absolute benchmark has an effective percentage of 96.9 percent, whereas the adjusted benchmark has an effective percentage of 64.7 percent.[1]

The above discussion addresses what the parties have described as the "granularity" problem with small sample benchmark application. That is, when failures are in integer increments, and the integer is larger than the permissible percentage of misses, then performance higher than the benchmark is required. The best illustration of this phenomenon is Example 1b above, where it is not possible to avoid identifying failures even though aggregate performance passes the benchmark. However, since benchmark adjustment can only be accomplished in integer increments, then adjustments can result in failing performance being identified as passing. This phenomenon is best illustrated in example 2b above, where no failures are identified even though aggregate performance fails the benchmark.

The granularity problem is distinguishable from a "random variation" problem. A random variation problem is illustrated in the following example:

---

[1] For example, there is approximately twice the number of results with a sample size of one as there is with a sample size of two (15.9 versus 8.7 percent, respectively; see column 6). Consequently, the average effective percentage across all sample sizes will be affected more by the effective percentage for sample sizes of one than by sample sizes of two. Table 1 columns 7 and 8 effectively account for this relative difference. This table is for illustration purposes only.

Example 3:

With ten orders each, altogether ten CLECs have 100 orders. Overall there are ten delinquent orders for a passing percentage of 90 percent. If each CLEC has one delinquent order, their performance results will all pass the benchmark. However, because of the "luck of the draw" it is unlikely that the delinquent orders will be distributed equally across all CLECs. Instead, two CLECs have two delinquent orders each, six CLECs have one delinquent order each, and two CLECs have no delinquent orders.

Thus, the performance results for two CLECs fail even though the overall performance passes the benchmark. Pacific has proposed statistical testing of benchmarks to mitigate the effects of random variation in this situation, which is analogous to the "Type I error" situation for parity measure assessment.

However, the parties have agreed to address the granularity problem, but not the random variation problem. The only exception is that the parties have agreed to use a statistical method to create the small sample adjustment tables. This appendix proceeds within these guidelines. A complete statistical treatment is not established. The granularity problem is addressed through statistically developed small sample adjustment tables.

Proposals

In summary, while small sample table allowances alleviate one problem, they cause another problem. Analyses of ILEC and CLEC proposals further illustrate this dilemma. The CLECs propose that benchmarks be strictly followed, causing instances where the ILECs must have 100 percent performance even though the parties have agreed that lower percentages allow a meaningful opportunity to compete. In compromise, the CLECs have offered small sample tables that allow performance thresholds to drop below the benchmark. However, these tables still may require overall performance levels to be well above the benchmark to avoid performance failure identification. For example, the CLEC-proposed small sample table for the 90 percent benchmark implies that the time criterion must be met

95.1 or 97.7 percent of the time for the ILEC to pass the measure– under conditions where the underlying process actually passes the benchmark.[2]

In contrast, the ILEC proposal for the 90 percent benchmark implies that the time criterion must be met 92.0 or 92.9 percent of the time for the ILEC to pass the measure[3] – again under conditions where the underlying process actually passes the benchmark. However both the ILEC and CLEC proposals' net result is to lower the effective benchmark performance level. Staff determined that the average effective level for the CLEC table was 89 percent and for the ILEC table was 83 percent.[4] For results with 100 or less orders, 13 and 27 percent of the results had effective benchmark percentages below 80 percent for the CLEC and ILEC small sample adjustment tables, respectively. The differences between CLEC and ILEC-proposed tables for the 95 and 99 percent benchmark parallel these differences.

There are a few aspects of small sample implementation that may allow maximizing the goals of both the ILECs and the CLECs, rather than trading the interests of one for the other. Staff examined four approaches:

---

[2] These percentages acknowledge random variation and assume a one-percent failure rate. The CLECs preferred using a derivation sample size of 20, which implies performance of 97.7 percent. However, the CLECs offered to compromise at a sample size of 100, which implies performance of 95.1 percent. Staff asked Pacific's Dr. Gleason for a copy of the MathCad© worksheet to calculate these implied performance levels. The worksheet is included as Attachment 1 to this appendix.

[3] This result also assumes only a one percent failure rate. Pacific prefers derivation sample sizes of 1000, which implies performance of 92.0 percent, but offers to use a derivation sample size of 400, which implies performance of 92.9 percent.

[4] These figures were calculated in several steps. First the effective performance level was calculated for each result in each table. For example, if for a sample of 4, one miss was allowed, the effective allowable performance level is 75 percent. Second, the percentage of sample sizes was calculated from January through May, 2000, data, for sample sizes of 100 or less (66% of total results). For example, samples with 2 orders accounted for 6.4 percent of these benchmark samples. Third, the effective percentage levels were weighted by these percentages. For example, since 7-order samples accounted for 3.2 percent of the results and 2-order samples accounted for 6.4 percent of the results, the 5-order samples' effective level was weighted twice the 7-order samples' level. And fourth, the overall average effective percentage was calculated from the weighted percentage. Using the same example, the 5-order samples' level essentially would be "counted" twice in determining the overall effective average compared to the 7-order samples' level. Example calculations were presented in Table 1.

(1) Application sample sizes currently are proposed to be the same for all benchmark percentage levels, and instead can be set to more closely fit the qualities of each different benchmark.
(2) Implied performance levels currently are proposed to require different degrees of improvement for different benchmarks; they can be made uniform by raising low improvement requirements and lowering high improvement requirements.
(3) Underlying process information: Aggregate CLEC results provide some indication of whether the underlying processes are passing or failing the benchmark, and thus can be used to guide a more targeted and appropriate application of the small sample tables.
(4) Small sample aggregation can alleviate the worst data "granularity" problems.

Application sample sizes
Applying small sample adjustment tables to the same sample size range for different benchmark percentage levels results in disparate treatment of the same problem. The "granularity" problem is of a different magnitude for different benchmark levels. For example, the sample size where a single delinquent order, or "miss," results in performance matching the nominal benchmark is very different for different benchmarks. For the 90 percent benchmark the sample size is 10. That is, one miss in ten equals 90 percent. In contrast, one miss in one hundred orders matches the 99 percent benchmark. Parallel to this phenomenon is the fact that the 99 percent benchmark experiences proportionately equal inaccuracies with much larger sample sizes than the 90 percent benchmark experiences. Or stated inversely, if applied to the same sample size range, the 99 percent benchmark experiences proportionately larger inaccuracies than the 90 percent benchmark. This results in a logical inconsistency where if the tolerances for one benchmark are optimized, then the other benchmarks are not optimal.

The solution described here selects reasonable tolerances and applies those tolerances consistently to all three benchmark percentage levels. Complete descriptions of the tolerances and methods to construct sample sizes is included as Attachment 2. Generally speaking, application sample sizes were selected that allowed discrepancies from the nominal benchmark percentage no greater than 10 percent of the allowable failure percentage.[5] This method resulted in sample sizes of 50, 100, and 500 for the 90, 95, and 99 percent benchmarks, respectively.[6]

Implied performance levels
Consistent implied performance criteria were used to select derivation sample sizes. First, the lower limit for each derivation sample size was determined by calculating the sample size that would allow no higher performance level than half the difference between the benchmark and 100 percent performance. For example, for the 90 percent benchmark, a derivation sample size of 125 implies a performance level of 95 percent, which is an increment of 5 percent or half the 10 percent allowed failures. Sample sizes were further adjusted by ensuring no effective adjusted percentage would be greater than the nominal benchmark percentage. This resulted in derivation sample sizes of 150, 300, and 1500 for the 90, 95, and 99 percent benchmarks, respectively. Attachment 2 lists the exact methods used.

Underlying process information
Application of small sample adjustment tables reduces what is analogous to a Type I error. That is, when an OSS process provides the service it should provide as defined by the benchmark, small sample adjustment tables reduce the likelihood of identifying spurious failures. On the other hand, adjustment tables increase the likelihood of what is analogous to a Type II error. That is, when an OSS process provides the service at levels lower than it should, small sample adjustment tables increase the likelihood that failures will not be detected.

Accuracy in benchmark decisions can be increased by applying small sample adjustment tables to situations where “Type I” errors are likely,

---

[5] For example, the 90 percent benchmark allows 10-percent failures. Ten percent of that failure allowance is one percent. Thus a small sample table is applied to all ranges of sample sizes where the average effective percentage is greater than 91 percent. Similarly, the 99 percent benchmark allows a one- percent failure. Ten percent of that failure allowance is one tenth of one percent. Thus a small sample table is applied to all ranges of sample sizes where the average effective percentage is greater than 99.1 percent. See Attachment 2 for the exact method used.

[6] Staff understands that pending Commission approval, only benchmark percentages of 90, 95, and 99 will be used in the final performance incentive plan. If it happens that other percentage levels are ultimately used, the method described in this appendix can be used to create any new sample size adjustment table.

and not to situations where "Type II" errors are likely. Analysis at the industry-wide aggregate performance level provides reasonably sufficient information regarding whether "Type I" or "Type II" errors are likely. If the aggregate performance level passes the benchmark, then CLEC-level performance that fails the benchmark has a greater likelihood of being a "Type I" error than if the aggregate performance failed the benchmark. And conversely, if the aggregate performance level fails the benchmark, then CLEC-level performance that passes the benchmark has a greater likelihood of being a Type II error than if the aggregate performance passed the benchmark. Therefore, a two-step small sample adjustment table application will maximize accuracy. First, since the large aggregate sample sizes provide a reasonable estimate of the process performance, they can be used to categorize performance results into "Type I error likely" and "Type II error likely" categories. For those samples where the aggregate performance passes ("Type I error likely"), small sample adjustment tables will be applied. For those samples where the aggregate performance fails ("Type II error likely"), small sample adjustment tables will not be applied – the nominal benchmark percentage will be the "pass/fail" criterion. Small sample adjustment tables will be used for the industry aggregate evaluation, however, in case any of these aggregate performance samples are small. The benefit of assessing and using the two categories to determine table application is best illustrated in the comparison of aggregate versus CLEC-specific sample sizes for the 90 and 99 percent benchmarks.[7] Table 2 shows the differences between aggregate and CLEC-specific result samples sizes. Aggregate sample sizes are typically large enough to assess whether "Type I" or "Type II" analogous error is likely for the much smaller CLEC-specific samples.

Table 2

Median sample sizes		
	Benchmark	
Sample type	90%	99%
CLEC-specific	36	5
Aggregate	9246	40725

[7] Sample sizes for the 95 percent benchmark are considerable larger. Only 14.7 percent of the CLEC-specific samples sizes are less than 100, and 14 percent of the aggregate sample sizes are less than approximately 9500.

Small sample aggregation

Table 1 illustrates that the granularity problem is most severe for the smallest samples. The most dramatic example is for the sample size of one. The only choices for an effective percentage level are zero and 100 percent. If the "ones" are aggregated, for example, into aggregates of five, then the choices are far better - 80 and 100 percent. Since aggregation of the smallest samples can alleviate the worst instances of the small sample problem, and since the equitable allocation of incentive payments still can be addressed in the incentive development phase of this proceeding, using the same aggregation rules as used for average-based parity measures can alleviate problems without unreasonably disadvantaging any party.

Methods

Based on the above principles, and using a MathCad© worksheet,[8] staff created small sample tables included here as Attachment 4. These tables are based on the following principles: (1) Application and derivation sample sizes should be set according to consistent relationships to the benchmark, (2) Implied performance should be no more than halfway between the benchmark and 100 percent performance.

Uniform criteria were applied to the three nominal benchmarks to establish the application sample size. These tables also provide a uniform limit of implied performance across the different benchmarks. They were constructed so that the implied performance would not exceed the midpoint between the benchmark nominal percentage and 100 performance. To accomplish this, the derivation sample sizes were 150, 300, and 1500 for the 90, 95, and 99 percent nominal benchmarks, respectively. With the midpoint expressed as 50 percent of the difference between the nominal percentage and 100 percent performance, the

---

[8] Staff requested Pacific Bell's consultant, Dr. Gleason, to provide staff with a copy of a program that would calculate the permitted misses for benchmarks in the form of a "small sample adjustment table." The Mathcad© program created by Dr. Gleason is included as Attachment 3. It is staff's understanding that AT&T's Dr. Mallows and Pacific's Dr. Gleason agreed on the methodology that Dr. Gleason subsequently forwarded to staff in the form of the MathCad© worksheets included here as Attachment 1 and 3. The worksheet in Attachment 1 calculates the implied performance level of different derivation sample sizes for different benchmark percentage levels. Staff used the worksheet to determine the sample sizes that would produce the desired performance level. Using the worksheet in Attachment 3, staff constructed the adjustment tables based on those sample sizes.

resultant implied performance difference was 44.2, 44.8, and 45.3 of the difference for the 90, 95, and 99 percent benchmarks, respectively.[9]

Summary

The small sample adjustment tables presented in Attachment 4 will be used in the following steps:

1. The number of performance "misses" for the CLEC industry-wide aggregate for each remedy plan benchmark submeasure will be compared to the number of permitted misses for all sample sizes covered by the related adjustment table. Industry aggregate performance will be identified as passing if the number of actual misses is less than or equal to the number of permitted misses, and identified as failing if otherwise.

2. For CLEC industry-wide aggregate sample sizes not covered by the related adjustment table, the actual performance percentage result will be compared to the benchmark nominal percentage value. Industry aggregate performance will be identified as passing if the actual performance percentage result is greater than or equal to the benchmark nominal percentage value, and identified as failing if otherwise.

3. For CLEC-specific analysis, results with sample sizes of four or less will be aggregated into a "small sample CLEC aggregate" for each submeasure. Each small sample CLEC aggregate performance result and all remaining non-aggregated CLEC performance results will be assessed.

4. For each submeasure where the CLEC industry-wide aggregate performance *fails* the benchmark, the actual performance percentage result for each small sample CLEC aggregate and each remaining non-aggregated CLEC result will be compared to the benchmark nominal

[9] In contrast, for example, the CLECs' percentage differences for the 3 benchmarks differed widely: 51.0, 63.7, and 85.1 for the 90, 95, and 99 percent benchmarks, respectively, using the derivation sample size of 100. Pacific's percentage differences for the 3 benchmarks also differed widely: 29.9, 40.3, and 67.9 for the 90, 95, and 99 percent benchmarks, respectively, using the derivation sample size of 400.

percentage value. Each individual or aggregate performance result will be identified as passing if the actual performance percentage result is greater than or equal to the benchmark nominal percentage value, and identified as failing if otherwise.

5. For sample sizes *covered* by the related adjustment table where the CLEC industry-wide aggregate performance *passes* the benchmark, the following shall apply for each submeasure. For each benchmark submeasure, the number of performance "misses" for each small sample CLEC aggregate and each remaining non-aggregated CLEC will be compared to the number of permitted misses. CLEC performance will be identified as passing if the number of actual misses is less than or equal to the number of permitted misses, and identified as failing if otherwise.

6. For sample sizes *not covered* by the related adjustment table where the CLEC industry-wide aggregate performance *passes* the benchmark, the following shall apply. The actual performance percentage result for each small sample CLEC aggregate and each remaining non-aggregated CLEC result will be compared to the benchmark nominal percentage value. Each individual or aggregate performance result will be identified as passing if the actual performance percentage result is greater than or equal to the benchmark nominal percentage value, and identified as failing if otherwise.

Mathcad worksheet: Small Sample Implied Performance Levels

Benchmarks for evaluation

$$B := (.9 \quad .95 \quad .97 \quad .99)^T$$

Reference sample sizes

$$N := (150 \quad 300 \quad 500 \quad 1500)^T$$

$i := 0 .. \text{length}(N) - 1$ $\qquad j := 0 .. \text{length}(B) - 1$

The following matrix gives the minimum number of "hits" consistent with the benchmark.

$$M_{i,j} := \text{ceil}\left(B_j \cdot N_i\right) \qquad M = \begin{bmatrix} 135 & 143 & 146 & 149 \\ 270 & 285 & 291 & 297 \\ 450 & 475 & 485 & 495 \\ 1.35 \cdot 10^3 & 1.425 \cdot 10^3 & 1.455 \cdot 10^3 & 1.485 \cdot 10^3 \end{bmatrix}$$

The following solve block calculates the performance level that meets the conditions that the failure rate on the benchmark given the sample size and the performance level should be 1%.

$$p_j := \frac{B_j + 1}{2}$$

Intial guesses for the solve block

Given

$$\text{pbinom}(M - 1, N, p) = .01$$

$$f(N, M, p) := \text{Find}(p)$$

$$X_{i,j} := f\left(N_i, M_{i,j}, p_j\right)$$

The matrix X has the performance levels by sample size and benchmark

$$X = \begin{bmatrix} 0.94416 & 0.98036 & 0.99139 & 0.99901 \\ 0.93371 & 0.97241 & 0.98612 & 0.99725 \\ 0.92721 & 0.96846 & 0.98352 & 0.99642 \\ 0.91666 & 0.96164 & 0.97877 & 0.99453 \end{bmatrix}$$

$Y := \text{augment}(N, X)$

$y := \text{augment}\left(0, B^T\right)$

$Z := \text{stack}(y, Y)$

The results are placed in a spreadsheet.

Z^T

Small sample adjustment table construction method.

To determine application sample sizes:

1. Calculate the net effective percentages for sample sizes without adjustments (i.e., use absolute application of benchmark percentage cutoffs).

2. Determine the table sub-ranges that are bounded by the different values where integer failures equal the benchmark. For example, for the 90-percent benchmark, the first integer failure results in a performance level equal to the benchmark at a sample size of 10. One (1) failure out of ten represents 90% performance. Thus the sub-ranges are 1-10, 21-30, 31-40, etc. For the 0.95 benchmark, the sub-range boundaries are 1-20, 21-40, 41-60, etc; and for the 0.99 benchmark the sub-range boundaries are 1-100, 101-200, 201-300, etc.

3. For each sub-range, exclude the sample size where the net effective percentage equals the benchmark and determine the average net effective percentage. For example, samples sizes of 10, 20, 30, etc., for the 90 percent benchmark are excluded. The application table will include all sub-ranges where the average effective percentage is greater than nominal benchmark value by 10 percent of the allowable “missed” percentage. For example, the 90-percent benchmark allows 10 percent “misses.” Ten percent of the allowable misses is one (1) percent. Therefore the table would include all sub-ranges where the average effective percentage is greater than 91 percent. The corresponding values for the 95 and 99-percent benchmarks are 95.5 and 99.1 percent, respectively. Following these criteria, adjustment tables will be applied to sample sizes of 50, 100, and 500 for the 90, 95, and 99 percent benchmarks, respectively.

Derivation sample sizes.

1. Smaller derivation sample sizes result in higher implied performance thresholds. Determine the lower bound for the derivation sample size by calculating adjustment tables that result in an implied performance value of no more than the midpoint of the interval between the nominal benchmark and perfect performance, 100 percent. For example, the midpoint between the 90-percent benchmark and 100 percent is 95 percent. Thus, the implied performance limits for adjustment tables are 95 percent for the 90 percent benchmark, 97.5 percent for the 95 percent benchmark, and 99.5 percent for the 99 percent benchmark.

2. Calculate the net effective benchmark percentages using adjustment tables to find the derivation sample size that is equal to or less than the limit. These values are 125, 246, and 1222 for the 90, 95, and 99 percent benchmarks, respectively.

3. If any net effective benchmark percentage is greater than the nominal benchmark percentage, then increase the derivation sample size until the net percentage is greater than the nominal percentage by no more than 10 percent of the allowable failure percentage. These individual sample size net effective percentage limits are 91, 95.5, and 99.1 percent for the 90, 95, and 99 percent nominal percentage benchmarks, respectively. Using these limits, the resulting derivation sample sizes are 150, 300, and 1500 for the 90, 95, and 99 percent nominal percentage benchmarks, respectively.

Mathcad worksheet: Creates small sample adjustment tables for different percentage benchmarks.

Set benchmarks for analysis.

$B := (.9 \quad .95 \quad .97 \quad .98 \quad .99 \quad .9925)^T$

Set reference sample size.

$N := 300$

Set probability of failing the benchmark at the reference sample size.

$P := .01$

Set probability of failing the benchmark with small samples.

$T1E := .1$

Set length of the small sample table.

$L := 100$

$J := \text{length}(B)$ $\qquad j := 0 .. J - 1$

The vector "b" gives the minimum number of successes permitted by the benchmark at the reference sample size.

$b_j := \text{ceil}\left(B_j \cdot N\right)$ $\qquad b^T = (270 \quad 285 \quad 291 \quad 294 \quad 297 \quad 298)$

The vector "p" gives initial guesses at the required performance levels

$$p_j := \frac{B_j + 1}{2}$$

The following function calculates performance levels that are consistent with the reference sample size N and probability P.

Given

$\text{pbinom}(b - 1, N, p) = P$

$f(b, p) := \mathrm{Find}(p)$

These are the required performance levels.

$F_j := f(b_j, p_j)$

$F^T = (0.9337126 \quad 0.9724129 \quad 0.9861193 \quad 0.9921846 \quad 0.9972458 \quad 0.9985427)$

Calculate the minimum number of misses for which the cumulative probability is less than T1E.

$$\mathrm{select}(n, P) := \left| \begin{array}{l} k \leftarrow 1 \\ \text{while } \ \mathrm{pbinom}(n - k, n, P) \geq T1E \\ \quad k \leftarrow k + 1 \\ \text{return } k - 1 \end{array} \right.$$

$k := 0 .. L - 1$

$K_k := k + 1$

$Y_{k,j} := \mathrm{select}(K_k, F_j)$

$Z := \mathrm{augment}(K, Y)$

$\mathrm{head}_{j+1} := B_j$

The following is the Small Sample Table.

$SST := \mathrm{stack}(\mathrm{head}^T, Z)$

Insert the Small Sample Table into an Excel spreadsheet

SST

Appendix K, Attachment 4

Benchmark Small Sample Adjustment Tables

90% Benchmark		95% Benchmark		99% Benchmark	
Sample size	Maximum permitted misses	Sample size	Maximum permitted misses	Sample size	Maximum permitted misses
1	0	1 to 3	0	1 to 19	0
2 to 9	1	4 to 19	1	20 to 97	1
10 to 20	2	20 to 40	2	98 to 202	2
21 to 31	3	41 to 63	3	203 to 319	3
32 to 44	4	64 to 88	4	320 to 445	4
45 to 50	5	89 to 100	5	446 to 500	5

Appendix L

List of Appearances

Respondents: Ed Kolto-Wininger and James B. Young, Attorneys at Law, for Pacific Bell; Marlin Ard and Elaine M. Duncan, Attorneys at Law, for Verizon California Inc.

Interested Parties: Evelyn C. Lee, Attorney at Law, for WorldCom, Inc.; Randolph Deutsch and Joseph Faber, Attorneys at Law, for AT&T Communications of California, Inc.; Richard L. Goldberg, Attorney at Law, for Sprint Communications Company LP; Theresa L. Cabral, Attorney at Law, for Mediaone Telecommunications of California and Karen Potkul, Attorney at Law, for XO, Inc. (formerly, Nextlink, Inc.)

Office of Ratepayer Advocates: Janice Grau, Attorney at Law.

(END OF APPENDIX L)