Carrier to Carrier Statistical Metric Evaluation Procedures

Statistical evaluation is used here as a tool to assess whether the Incumbent Local Exchange Company's (ILEC) wholesale service performance to the Competitive Local Exchange Companies (CLECs) is at least equal in quality to the service performance that the ILEC provides to itself (i.e., parity). Carrier-to-Carrier (C2C) measurements having a parity standard are metrics where both the CLEC and ILEC performance are reported.[1]

A. Statistical Framework

The statistical tests of the null hypothesis of parity against the alternative hypothesis of non-parity defined in these guidelines use ILEC and CLEC observational data. The ILEC and CLEC observations for each month are treated as random samples drawn from operational processes that run over multiple months. The null hypothesis is that the CLEC mean performance is at least equal to or better than the ILEC mean performance.

Statistical tests should be performed under the following conditions.

1) The data must be reasonably free of measurement/reporting error.

2) The ILEC to CLEC comparisons should be reasonably like to like.

3) The minimum sample size requirement for statistical testing is met. (Section B)

4) The observations are independent. (Section D)

These conditions are presumed to be met until contrary evidence indicates otherwise.

To the extent that the data and/or operational analysis indicate that additional analysis is warranted, a metric may be taken to the Carrier Working Group for investigation.

B. Sample Size Requirements

[1] Section 251(c)(2)(C) of the Telecommunications Act of 1996 states that facilities should be provided to CLECs on a basis "that is at least equal in quality to that provided by the local exchange carrier to itself." Paragraph 3 of Appendix B of FCC Opinion 99-404 states, "Statistical tests can be used as a tool in determining whether a difference in the measured values of two metrics means that the metrics probably measure two different processes, or instead that the two measurements are likely to have been produced by the same process."

The assumptions that underlie the C2C Guidelines statistical models include the requirement that the two groups of data are comparable. With larger sample sizes, differences in characteristics associated with individual customers are more likely to average out. With smaller sample sizes, the characteristics of the sample may not reasonably represent those of the population. Meaningful statistical analysis may be performed and confident conclusions may be drawn, if the sample size is sufficiently large to minimize the violations of the assumptions underlying the statistical model.

The following sample size requirements, based upon both statistical considerations and also some practical judgment, indicate the minimum sample sizes above which parity metric test results (for both counted and measured variables) may permit reasonable statistical conclusions.

The statistical tests defined in these guidelines are valid under the following conditions:

If there are only 6 of one group (ILEC or CLEC), the other must be at least 30.
If there are only 7 of one, the other must be at least 18.
If there are only 8 of one, the other must be at least 14.
If there are only 9 of one, the other must be at least 12.
Any sample of at least 10 of one and at least 10 of the other is to be used for statistical evaluation.

When a parity metric comparison does not meet the above sample size criteria, it may be taken to the Carrier Working Group for alternative evaluation. In such instances, a statistical score (Z score equivalent) will not be reported, but rather an "SS" (for Small Sample) will be recorded in the statistical score column; however, the means (or proportions), number of observations and standard deviations (for means only) will be reported.

C. Statistical Testing Procedures

Parity metric measurements that meet the sample size criteria in Section B will be evaluated according to the one-tailed permutation test procedure defined below.

Combine the ILEC and CLEC observations into one group, where the total number of observations is $n_{ilec+}\, n_{clec}$. Take a sufficiently large number of random samples of size n_{clec} (e.g., 500,000). Record the mean of each re-sample of size n_{clec}. Sort the re-sampled means from best to worst (left to right) and compare where on the distribution of re-sampled means the original CLEC mean is located. If 5% or less of the means lie to the right of the reported CLEC mean, then reject the null hypothesis that the original CLEC sample and the original ILEC sample came from the same population.

If the null hypothesis is correct, a permutation test yields a probability value (*p value)* representing the probability that the difference (or larger) in the ILEC and CLEC sample means is due to random variation.

Permutation test *p values* are transformed into "Z score equivalents." These "Z score equivalents" refer to the standard normal Z score that has the same probability as the p-values from the permutation test. Specifically, this statistical score equivalent refers to the inverse of the standard normal cumulative distribution associated with the probability of seeing the reported CLEC mean, or worse, in the distribution of re-sampled permutation test means. A Z score of less than or equal to –1.645 occurs at most 5% of the time under the null hypothesis that the CLEC mean is at least equal to or better than the ILEC mean. A Z score greater than –1.645 (p-value greater than 5%) supports the belief that the CLEC mean is at least equal to or better than the ILEC mean. For reporting purposes, Z score equivalents equal to or greater than 5.0000 are displayed on monthly reports as 5.0000. Similarly, values for a Z statistics equal to or less than –5.0000 are displayed as –5.0000.

Alternative computational procedures (i.e., computationally more efficient procedures) may be used to perform measured and counted variable permutation tests so long as those procedures produce the same p-values as would be obtained by the permutation test procedure described above. The results should not vary at or before the fourth decimal place to the Z score equivalent associated with the result generated from the exact permutation test. (i.e., the test based upon the exact number of combinations of n_{clec} from the combined n_{ilec+} n_{clec}).

Measured Variables (i.e., metrics of intervals, such as mean time to repair or average delay days):

The following permutation test procedure is applied to measured variable metrics:

1. Compute and store the mean for the original CLEC data set.
2. Combine the ILEC and CLEC data to form one data set.
3. Draw a random sample without replacement of size n_{clec} (sample size of original CLEC data) from the combined data set.

   a) Compute the test statistic (re-sampled CLEC mean).
   b) Store the new value of test statistic for comparison with the value obtained from the original observations.
   c) Recombine the data set.

4. Repeat Step 3 enough times such that if the test were re-run many times the results would not vary at or before the fourth decimal place of the reported Z score equivalent (e.g., draw 500,000 re-samples per Step 3).
5. Sort the CLEC means created and stored in Step 3 and Step 4 in ascending order (CLEC means from best to worst).

6. Determine where the original CLEC sample mean is located relative to the collection of re-sampled CLEC sample means. Specifically, compute the percentile of the original CLEC sample mean.
7. Reject the null hypothesis if the percentile of the test statistic (original CLEC mean) for the observations is less than .05 (5%). That is, if 95% or more of the re-sampled CLEC means are better than the original CLEC sample mean, then reject the null hypothesis that the CLEC mean is at least equal to or better than the ILEC mean. Otherwise, the data support the belief that the CLEC mean is at least equal to or better than the ILEC mean.
8. Generate the C2C Report "Z Score Equivalent," known in this document as the standard normal Z score that has the same percentile as the test statistic.

Counted Variables (i.e., metrics of proportions, such as percent measures):

A hypergeometric distribution based procedure (a.k.a., Fisher's Exact test)[2] is an appropriate method to evaluate performance for counted metrics where performance is measured in terms of success and failure. Using sample data, the hypergeometric distribution estimates the probability (*p value*) of seeing **at least** the number of failures found in the CLEC sample. In turn, this probability is converted to a Z score equivalent using the inverse of the standard normal cumulative distribution.

The hypergeometric distribution is as follows:

$$p\ value = 1 - \left\{ \sum_{i=\max(0,\{[n_{ilec}p_{ilec}+n_{clec}p_{clec}]+[n_{clec}]-[n_{ilec}+n_{clec}]\})}^{n_{clec}p_{clec}-1} \frac{\binom{[n_{clec}p_{clec}+n_{ilec}p_{ilec}]}{i}\binom{[n_{clec}+n_{ilec}]-[n_{clec}p_{clec}+n_{ilec}p_{ilec}]}{n_{clec}-i}}{\binom{[n_{clec}+n_{ilec}]}{n_{clec}}} \right\}$$

Where:

p value = the probability that the difference in the ILEC and CLEC sample proportions could have arisen from random variation, assuming the null hypothesis

n_{clec} and n_{ilec} = the CLEC and ILEC sample sizes (i.e., number of failures + number of successes)

p_{clec} and p_{ilec} = the proportions of CLEC and ILEC failed performance, for percentages 10% translates to a 0.10 proportion = number of failures / (number of failures + number of successes)

---

[2] This procedure produces the same results as a permutation test of the equality of the means for the ILEC and CLEC distributions of 1s and 0s, where successes are recorded as 0s and failures as 1s.

Either of the following two equations can be used to implement a hypergeometric distribution-based procedure:

The probability of observing **exactly** f_{clec} failures is given by:

$$\Pr(i = f_{clec}) = \frac{\binom{(f_{clec} + f_{ilec})}{f_{clec}} \binom{(n_{clec} + n_{ilec}) - (f_{clec} + f_{ilec})}{n_{clec} - f_{clec}}}{\binom{(n_{clec} + n_{ilec})}{n_{clec}}}$$

Where:

f_{clec} = CLEC failures in the chosen sample = $n_{clec}\, p_{clec}$
f_{ilec} = ILEC failures in the chosen sample = $n_{ilec}\, p_{ilec}$
n_{clec}= size of the CLEC sample
n_{ilec}= size of the ILEC sample

Alternatively, the probability of observing **exactly** f_{clec} failures is given by:

$$\Pr(i = f_{clec}) = \frac{n_{clec}!\, n_{ilec}!\, f_{total}!\, s_{total}!}{(n_{clec} + n_{ilec})!\, f_{clec}!\, (n_{clec} - f_{clec})!\, (f_{total} - f_{clec})!\, (n_{ilec} - f_{total} + f_{clec})!}$$

Where:

s_{clec} = the number of CLEC successes = $n_{clec}\, (1-p_{clec})$
s_{ilec} = the number of ILEC successes = $n_{ilec}\, (1-p_{ilec})$
$f_{total} \equiv f_{clec} + f_{ilec}$
$s_{total} \equiv s_{clec} + s_{ilec}$

The probability of observing f_{clec} **or more** failures [$Pr(i \geq f_{clec})$] is calculated according to the following steps:

1. Calculate the probability of observing exactly f_{clec} using either of the equations above.
2. Calculate the probability of observing all more extreme frequencies than $i = f_{clec}$, conditional on the
    a. total number of successes (s_{total}),
    b. total number of failures (f_{total}),
    c. total number of CLEC observations (n_{clec}), and the
    d. total number of ILEC observations (n_{ilec}) remaining fixed.

3. Sum up all of the probabilities for $Pr(i \geq f_{clec})$.
4. If that value is less than or equal to 0.05, then the null hypothesis is rejected.

D. Root Cause/Exceptions

Root Cause: If the permutation test shows an “out-of-parity” condition, the ILEC may perform a root cause analysis to determine cause. Alternatively, the ILEC may be required by the Carrier Working Group to perform a root cause analysis. If the cause is the result of “clustering” within the data, the ILEC will provide such documentation.

Clustering Exceptions: Due to the definitional nature of the variables used in the performance measures, some comparisons may not meet the requirements for statistical testing. Individual data points may not be independent. The primary example of such non-independence is a cable failure. If a particular CLEC has fewer than 30 troubles and all are within the same cable failure with long duration, the performance will appear out of parity. However, for all troubles, including the ILEC’s troubles, within that individual event, the trouble duration is identical.

Another example of clustering is if a CLEC has a small number of orders in a single location with a facility problem. If this facility problem exists for all customers served by that cable and is longer than the average facility problem, the orders are not independent and clustering occurs.

Finally, if root cause shows that the difference in performance is the result of CLEC behavior, the ILEC will identify such behavior and work with the respective CLEC on corrective action.

Another assumption underlying the statistical models used here is the assumption that the data are independent. In some instances, events included in the performance measures of provisioning and maintenance of telecommunication services are not independent. The lack of independence contributes to “clustering” of data. Clustering occurs when individual items (orders, troubles, etc.) are clustered together as one single event. This being the case, the ILEC will have the right to file an exception to the performance scores in the Performance Assurance Plan if the following events occur:

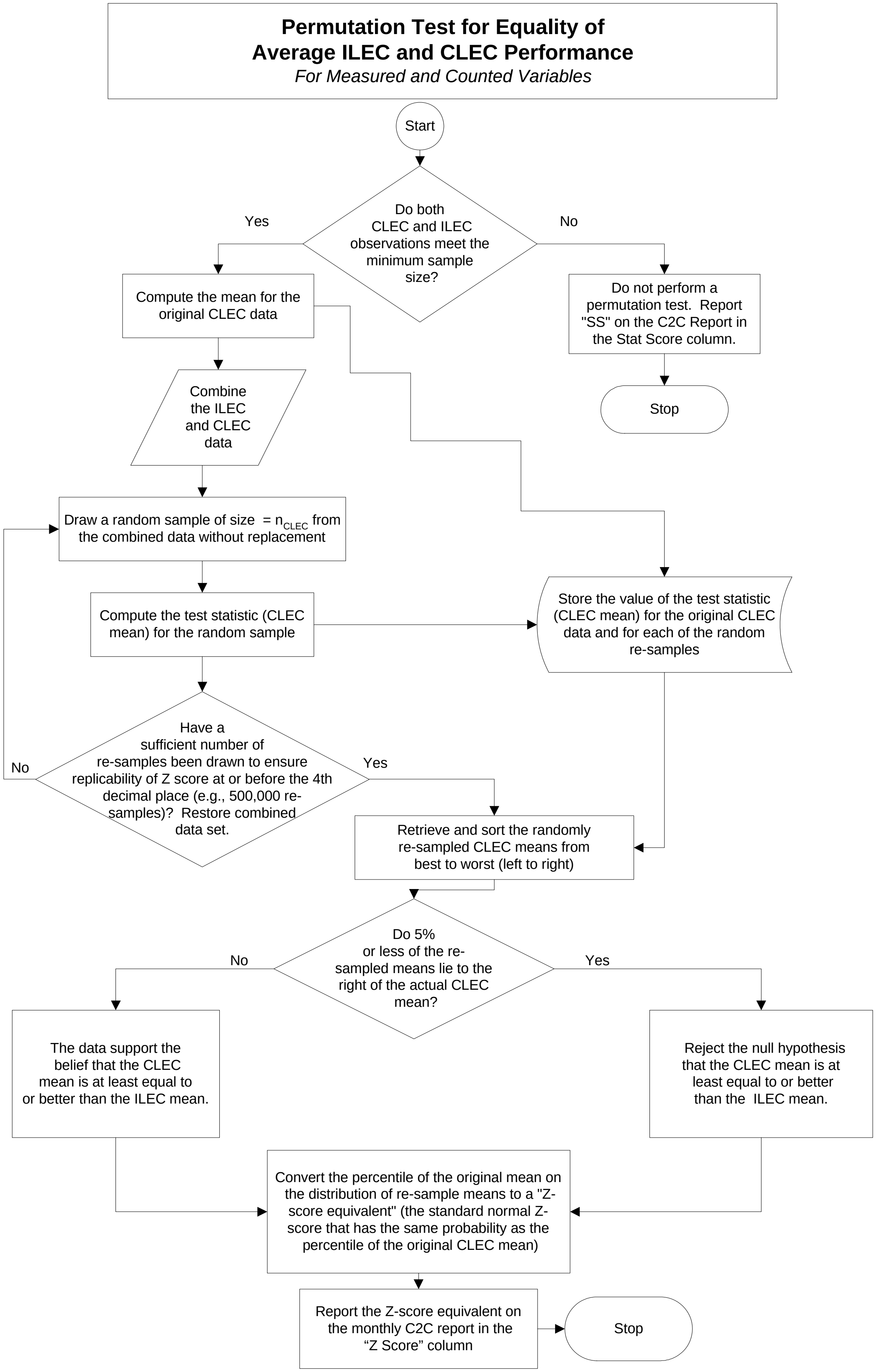
a. Event-Driven Clustering - Cable Failure: If a significant proportion (more than 30%) of a CLEC’s troubles are in a single cable failure, the ILEC may provide data demonstrating that all troubles within that failure, including the ILEC troubles, were resolved in an equivalent manner. Then, the ILEC also will provide the repair performance data with that cable failure performance excluded from the overall performance for both the CLEC and the ILEC and the remaining troubles will be compared according to normal statistical methodologies.

b. Location-Driven Clustering - Facility Problems: If a significant proportion (more than 30%) of a CLEC's missed installation orders and resulting delay days were due to an individual location with a significant facility problem, the ILEC will provide the data demonstrating that the orders were "clustered" in a single facility shortfall. Then, the ILEC will provide the provisioning performance with that data excluded. Additional location-driven clustering may be demonstrated by disaggregating performance into smaller geographic areas.

c. Time-Driven Clustering - Single Day Events: If a significant proportion (more than 30%) of CLEC activity, provisioning, or maintenance occurs on a single day within a month, and that day represents an unusual amount of activity in a single day, the ILEC will provide the data demonstrating the activity is on that day. The ILEC will compare that single day's performance for the CLEC to the ILEC's own performance. Then, the ILEC will provide data with that day excluded from overall performance to demonstrate "parity."

CLEC Actions: If performance for any measure is impacted by unusual CLEC behavior, the ILEC will bring such behavior to the attention of the CLEC to attempt resolution. Examples of CLEC behavior impacting performance results include order quality, causing excessive missed appointments; incorrect dispatch identification, resulting in excessive multiple dispatch and repeat reports, inappropriate X coding on orders, where extended due dates are desired; and delays in rescheduling appointments, when the ILEC has missed an appointment. If such action negatively impacts performance, the ILEC will provide appropriate detailed documentation of the events and communication to the individual CLEC and the Commission.

Documentation: The ILEC will provide all necessary detailed documentation to support its claim that an exception is warranted, ensuring protection of customer proprietary information, to the CLEC(s) and Commission. ILEC and CLEC performance details include information on individual trouble reports or orders. For cable failures, the ILEC will provide appropriate documentation detailing all other troubles associated with that cable failure.


Permutation Test for Equality of
Average ILEC and CLEC Performance
For Measured and Counted Variables
Start
Do both CLEC and ILEC observations meet the minimum sample size?
Yes
No
Compute the mean for the original CLEC data
Do not perform a permutation test. Report "SS" on the C2C Report in the Stat Score column.
Stop
Combine the ILEC and CLEC data
Draw a random sample of size = n_{CLEC} from the combined data without replacement
Compute the test statistic (CLEC mean) for the random sample
Store the value of the test statistic (CLEC mean) for the original CLEC data and for each of the random re-samples
Have a sufficient number of re-samples been drawn to ensure replicability of Z score at or before the 4th decimal place (e.g., 500,000 re-samples)? Restore combined data set.
No
Yes
Retrieve and sort the randomly re-sampled CLEC means from best to worst (left to right)
Do 5% or less of the re-sampled means lie to the right of the actual CLEC mean?
No
Yes
The data support the belief that the CLEC mean is at least equal to or better than the ILEC mean.
Reject the null hypothesis that the CLEC mean is at least equal to or better than the ILEC mean.
Convert the percentile of the original mean on the distribution of re-sample means to a "Z-score equivalent" (the standard normal Z-score that has the same probability as the percentile of the original CLEC mean)
Report the Z-score equivalent on the monthly C2C report in the "Z Score" column
Stop

Statistical Methodologies:

For performance measures where "parity" is the standard, Verizon will use the following tests:

Sample Sizes	Means:	Proportions:	Rates:
"Large samples"	Modified t $t = \frac{\bar{X}_{clec} - \bar{X}_{vz}}{\sqrt{s_{vz}^2\left(\frac{1}{n_{vz}} + \frac{1}{n_{clec}}\right)}}$	Modified t $t = \frac{p_{clec} - p_{vz}}{\sqrt{p_{vz}(1-p_{vz})\left(\frac{1}{n_{vz}} + \frac{1}{n_{clec}}\right)}}$	Modified Z $Z = \frac{r_{clec} - r_{VZ}}{\sqrt{r_{vz}\left(\frac{1}{b_{vz}} + \frac{1}{b_{clec}}\right)}}$
"Small samples"	Permutation testing	Fisher's exact test	Binomial exact test

Note: If the metric is one where a higher mean, proportion or rate signifies better performance, the means, proportions, or rates in the numerator of the statistical formulas should be reversed.

Definitions:

$\bar{X}_i$ is the sample mean where *i = CLEC, VZ.*

p_i is the sample proportion where $0.000 < p_i < 1.000$ and where *i = CLEC,VZ*.

r_i is the sample rate where *i = CLEC, VZ.*

s_{vz}^2 is the sample VZ variance.

n_i is the number of transactions where *i = CLEC, VZ.*

n is the total number of transactions ($\sum_1^i n_i$).

b_i is the number of base elements where *i = CLEC, VZ.*

b is the total number of base elements ($\sum_1^i b_i$).

q_{vz} is the relative proportion of base elements such that $q_{vz} = \frac{b_{vz}}{b}$.

Procedures for testing differences between CLEC and Verizon performance

1. If the CLEC performance is better than or equal to the Verizon performance, no testing will be done.
2. If the CLEC performance is worse than the Verizon performance,
   a. For means: If $n_i \geq 30$, the modified t-test will be used. If $n_i < 30$, the modified t-test will be used until permutation testing can be done in an automated fashion.
   b. For proportions: If $n_i p_i (1 - p_i) \geq 5$, the modified t-test will be used. Otherwise Fisher's exact test will be used.

c. For rates: Until the binomial test can be run for all samples in an automated fashion, the following sample size condition will apply: If $nq_{vz}\left(1-q_{vz}\right)\geq 5$, the modified Z-test described above will be used Otherwise, the binomial test (non-automated) will be used.