AT&T

Rhonda P. Merritt
Law & Government Affairs
Assistant Vice President - Florida

Suite 700
101 N. Monroe St.
Tallahassee, FL 32301
850 425-6342
FAX: 850 425-6361

ORIGINAL

December 20, 2000

RECEIVED-FPSC
00 DEC 20 PM 2:37
RECORDS AND REPORTING

Ms. Blanca Bayo, Director
Division of Records and Reporting
Florida Public Service Commission
2540 Shumard Oak Boulevard
Tallahassee, Florida 32399-0850

RE: Docket 000121-TP; Investigation into the establishment of operations support systems permanent performance measures for incumbent local exchange telecommunications companies.

Dear Ms. Bayo:

Attached is AT&T's filing of the following items in the above docket:

- Joint CLEC Performance Incentive Plan
- CLEC Proposed Disaggregation (Process Level)
- Measure Concentrated CLECs and the Maximum Balancing Critical Value

All of these documents have been provided in electronic format to Mr. Paul Stallcup.

Sincerely,

Rhonda Merritt

Rhonda P. Merritt

APP
CAF
CMP 3
COM
CTR
ECR 1
LEG 1
OPC
PAI
RGO 1 Harvey
SEC 1
SER
OTH Conn Allen

cc: Paul Staffcup

RECEIVED & FILED
mw
FPSC-BUREAU OF RECORDS

DOCUMENT NUMBER-DATE
16287 DEC 20 8
FPSC-RECORDS/REPORTING

Joint CLEC

Performance Incentive Plan

Introduction

It is well recognized that a meaningful system of self-enforcing consequences for discriminatory ILEC performance is critically important to the protection of the public's interest and the rapid and sustainable development of a competitive local telecommunications market. Incumbent LECs have strong business incentives and means to maintain their current monopolies through the delivery of inadequate and unlawful levels of operations support for CLECs. Thus, an appropriate system of self-enforcing consequences is absolutely necessary to assure that the competitive local telecommunications markets envisioned by the 1996 Act will be able to develop and survive.

In order to be effective, prompt enforcement of appropriate consequences must be assured. Because of the extensive delays inherent in the adjudication and appeals process, CLECs cannot rely solely upon the legal/regulatory process to obtain appropriate remedies for discriminatory ILEC performance. Furthermore, the consequences must provide ILECs with incentives that exceed the benefits it may derive by inhibiting competition, and such consequences must be immediately imposed upon a demonstration of poor ILEC performance. The objective is to set the incentives in amounts that encourage ILECs to take proactive steps to prevent its performance from becoming non-compliant and, when it does reach that level, to correct its performance failures promptly.

It is beyond dispute that any system of self-enforcing consequences must be based upon an underlying set of performance measurements that cover the full panoply of ILEC activities upon which CLECs must rely to deliver their own retail service offerings. The Act requires that these activities, which touch upon every aspect of the business relationship between incumbents and CLECs, must be provided in a non-discriminatory manner. Thus, the interconnection agreements between incumbents and CLECs should ideally serve as a source for performance measurements. However, experience in Florida and elsewhere has proven that CLECs have generally been unable to individually negotiate, or even arbitrate, a sufficiently robust set of performance measurements.[1] For that reason, the first step in constructing a system of self-enforcing consequences must include careful consideration of the adequacy of the underlying measurement set. At a minimum, the performance measurements must supply each CLEC with reliable data on the incumbent's performance for that CLEC. Such data must be sufficiently discrete (as to the processes monitored) and detailed (to isolate and compare only comparable conditions) so as to permit a CLEC to enforce the terms of its interconnection agreement with the incumbent. In addition, the underlying performance measurement system should demonstrate quality implementation of the following characteristics:

— A comprehensive set of comparative measurements that monitors all areas of support (i.e., pre-ordering, ordering, provisioning, maintenance & repair and billing) without preference to any particular mode of market entry

— Measurements and methodologies that are documented in detail so that clarity exists regarding what will be measured, how it will

[1] As a starting point, the CLEC industry generally supports the measurement areas as documented in Local Competition Users Group (LCUG) – Service Quality Measurements

be measured and in what situations a particular event may be excluded from monitoring (such exclusions must also be tracked and reported)

- Sufficient disaggregation of results, so that only the results for similar operational conditions are compared and, particularly, so that the averaging of results will not mask discrimination[2]

- Pre-specified and pro-competitive performance standards exist. This includes identifying reasonably analogous performance delivered by the incumbent to its own operations[3] or, when such comparative standards are not readily identifiable, then absolute minimum standards for performance (benchmarks) are established[4]

- Sound quantitative methodology is used to compare CLEC experiences to analogous incumbent support[5]

- The overall performance measurement system is subject to initial and periodic validation, in order to assure that the performance

(SQMs), Version 7.0, August 28, 1998.

[2] The importance of sufficient disaggregation is more fully discussed in Attachment A

[3] Analogous performance must be broadly interpreted and consider not only retail operations of the incumbent but also operations of affiliates. Often the incumbent's asserted lack of analogous performance relies upon very narrow (and inappropriate) interpretation of the term "analogous" to mean "precisely identical" rather than "similar in key aspects." Furthermore, if the incumbent delivers different levels of performance to an affiliate and its the retail operations, the CLEC experience should be compared to the better of the two.

[4] In all cases, benchmarks must provide an efficient competitor with a meaningful opportunity to compete.

[5] As a general rule, when benchmarks are employed, statistical comparisons of the measured result for the CLEC to the benchmark are not appropriate. Typically, the standards state a minimum performance level that is required to support effective competition and the minimum success level that must be demonstrated to attain the benchmark. Thus, the typical form of the standard is, for example, "95% installed within 3 days." Note that in the preceding example a 5% deviation from the benchmark is permitted and, as a result, the potential for random variation of the performance is fully addressed. Any further accommodation of variation, as would occur if statistical procedures were employed, would effectively "double count" forgiveness of variability.

results which form the foundation for all decisions regarding the quality of the performance delivered by the ILEC are correct representations of the CLECs' marketplace experience.

It is critical that a performance measurement system incorporating all of the above characteristics exist before applying an incentive plan, because a robust and independently audited performance measurement system is a prerequisite to any effective system of self-enforcing consequences.[6]

Objectives of the Plan

A system of self-enforcing consequences must fully implement the following objectives:

- Consequences must be based upon the quality of support delivered on individual measures to individual CLECs
- Total consequences, in the aggregate, must have sufficient impact to motivate compliant performance without the need to apply a remedy repeatedly
- The imposition of financial consequences must be prompt and certain, and consequences should be self-executing so that opportunities for delay through litigation and regulatory review are minimized

[6] For example, business rules for individual performance measurements may provide for automatic exclusions of data points from the calculation. If such provisions are made, however, the exclusions must be according to clearly defined rules and the number of data points excluded for each submeasurement and for each CLEC should be reported on a monthly basis.

- Consequences must escalate as the basis for concluding that a performance failure exists becomes more substantial and/or the performance repeatedly fails to meet the applicable standard
- Additional consequences must apply when non-compliant performance is provided to CLECs on an industry-wide basis
- Exclusions from consequences must be minimized and the exclusions that are provided for must be monitored and limited to assure they do not mask discrimination
- Incumbents must have minimal opportunities to avoid consequences through such means as liability caps, offsetting credits, or a requirement that CLECs must demonstrate an ILEC's intent to harm
- Potential "entanglement" costs must be minimized so that, for example, access to mitigation measures for the incumbent does not become a means to revert to the legal/regulatory process and delay the application of consequences that should be self-enforcing

Structure of Consequences for Discriminatory ILEC Performance

Consequences operating on two tiers are proposed. The first tier addresses the consequences for non-compliant performance delivered to an individual CLEC. The second addresses the consequences for non-compliant performance delivered to the CLEC industry as a whole. In general terms, Tier I provides a form of non-exclusive liquidated damages payable to individual CLECs. Tier II, by contrast, incorporates what can be characterized as regulatory fines that are necessary when the ILEC's performance affects the competitive market – and consumers -- as a whole.

The total amount of Tier I payments (which are only an estimate of the CLECs' actual damages) is unlikely to provide the ILEC with sufficient incentives to take the actions necessary to eliminate its monopoly. Rather, an ILEC may decide to treat such payments as the price for retaining its monopoly and voluntarily incur them as a cost of doing business. Moreover, the harm that results when the ILEC provides discriminatory support for the CLEC industry in the aggregate has a major impact not only on CLECs but also on the operation of the competitive marketplace in general, which directly affects all Florida consumers of telecommunications services. Thus, it is appropriate to establish incentives to prevent this type of harm from occurring (or continuing), and both Tier I and Tier II are necessary and complementary elements of an effective system of consequences. Together, they work in tandem to achieve the goals of the Act.

Tier I

A Tier I consequence should be payable to an affected CLEC whenever any performance result indicates support delivered by the ILEC to an individual CLEC fails to meet or exceed the applicable performance standard.[7]

The first step in establishing Tier I consequences is to define the rule for determining if performance for a particular period "passes" or "fails" and, if it fails, whether additional consequences are warranted. Defining "pass/fail" rules requires that the underlying measurements be mapped into one of two classes:

[7] In the course of establishing Tier I consequences, the rights of an individual CLEC to pursue actual damages must be retained. However, if a CLEC sought to pursue a claim for actual damages, it would be reasonable to offset the damage award by any Tier I payments it received from the ILEC for the same time period and performance areas. In addition, a CLEC must retain the right to waive Tier I claims and pursue its individually negotiated contract remedies (if and only if the claims and remedies are not mutually payable.).

(1) those for which the performance standard is parity with analogous incumbent LEC performance results, and

(2) those for which the performance standard is an absolute level of required performance (otherwise known as a benchmark)

The differentiation is important because when parity is the standard, statistical procedures are usually necessary to draw conclusions regarding compliance. In such situations (which should apply to the vast majority of cases), two separate data sets are compared – one for the CLEC and one for the ILEC. Each data set is characterized by a mean and standard deviation. Statistical tests are used to draw a conclusion regarding the likelihood that the data sets with the observed means and standard deviations were drawn from the same population (in this case a support process for CLECs with the same quality and/or timeliness as that employed for the ILEC). The proper test further allows determination that parity does not exist, but it does not quantify "how far out of parity" the process is when parity is not indicated.[8]

In contrast, when a benchmark serves as the performance standard, measurement establishes a performance failure directly and assesses the degree to which performance departs from the standard. As explained below, the detailed mechanism for determining a performance failure differs for each of these types of measurement standards, but the principle governing the application of the Tier I consequence is consistent: the consequence escalates with increasing evidence and level of non-compliant performance.

[8] Clearly, however, when all other factor are held constant, increased statistical confidence is directly correlated (monotonic) with larger differences in the two sample

Tier I Business Rules for Parity Measurements

1. Use the Modified z-Statistic to Determine Compliance

The determination of whether performance is compliant (i.e., equal to or better than the appropriate standard) is based on the calculation of the modified z-statistic (z).[9] The calculated modified z-statistic is then compared to the cumulative normal distribution table to determine if parity exists.[10] For any such decision rule, the probability of an erroneous decision is known. For example, if the critical value is −3.00 and parity actually exists, the probability of saying it is not is 0.13%.

2. Use Permutation Analysis for Small Samples

Permutation analysis is employed for small data sets (those with 30 or fewer observations in one of the data sets to be compared) to create a probability distribution as an alternative to the cumulative normal distribution.[11] By

---

means being compared and therefore is a reasonable indication of how different ILEC performance was for itself versus that of the CLEC in the period of observation.

[9] See: Local Competition Users Group - Statistical Tests for Local Service Parity, February 6, 1998, Version 1.0 for documentation of the calculation and use of the modified z-statistic, which is included as Attachment B.

[10] The modified z-statistic computation provides for the CLEC mean to be subtracted from the ILEC mean. Thus, a negative z-statistic critical value presumes that worse performance exists when the CLEC mean becomes larger than the ILEC mean. For example, worse performance exists when the order completion interval for the CLEC exceeds that for the ILEC. Thus a negative z-statistic critical value is appropriate. On the other hand, for a metric like "% completed within x days", worse performance for the CLEC occurs when the metric result is smaller for the CLEC vis-à-vis the ILEC. In this case a positive z-statistic critical value is appropriate.

[11] See Attachment C for a description of the procedural steps for performing permutation analysis. Again, BST and the CLECs generally concur that permutation analysis is appropriate for data sets of this size.

mutual agreement, permutation analysis can also be employed for larger data sets.

3. Use the Balancing Critical Value

The threshold level to determine whether or not a performance failure exists is established by balancing Type I and Type II error.[12] This balance point is a function of the size of the CLEC data set (assuming the ILEC data set is very large) and the extent to which the means for the two data sets differ (assuming that both data sets are normally distributed). Simulation comparing relatively small data sets (as would be likely for a CLEC) to a much larger data set (as would likely exist for an ILEC) demonstrates that the balancing of Type I and Type II error can reasonably be expected to occur in the range of 25% for "samples" with fewer than 100 data points but is about 5% for samples with 1000 data points.[13] An appropriate method for calculating the critical values which depend on the sample size and balances Type I and Type II error probabilities for each given submeasure is specified in Attachment G.

The Joint CLEC PIP proposes a floor BCV of –3.1 implying a .001 level of significance (probability of Type I error) for the test. In other words, for a given delta, the BCV approach is employed as sample size increases until the probability of a Type I (=probability of Type II) error falls to .001, i.e., until the BCV reaches –3.1, and then "balancing is stopped." As sample size

[12]The key consideration is balancing the probability of drawing erroneous conclusions -- either that performance is "bad" when it is actually "good" (Type I error) or that performance is "good" when it is actually "bad" (Type II error). The former error adversely impacts ILECs and the latter adversely impacts CLECs. Unfortunately, reducing the likelihood of one type of error increases the likelihood of the other type of error occurring. Thus the best means to create an equitable outcome for all parties is to balance the Type I and Type II error.

increases beyond this point, the Type II error falls, but the probability of Type I error remains fixed at .001.[14] At this level of significance, the difference between Type I and Type II can be no more than 0.1%, which for all practical purposes implies the errors are balanced. Further, the .001 significance level is large enough to alleviate the testing impact of Type I and Type II errors – i.e., the payment or non-payment of penalties due to false positives or negatives.

Furthermore, the definition of the alternative hypothesis required to perform the balancing is fundamental to the applicability of the method. The Joint CLEC PIP proposes a value of 0.25 or less for the parameter δ. .[15][16]

4. Increase Consequences as the Confidence in a "Non-Parity" Conclusion Increases

An appropriate means to take increased confidence into consideration is to provide for higher amounts of monetary consequences as the confidence in the "non-parity" conclusion increases. This is justified because (all other factors held constant) as the difference in the mean performance for the

[13] See Response to Question 3 contained in AT&T Ex Parte filed in CC Docket 98-56 dated July 13, 1999.

[14] Even as sample size drives the probability of Type I error to zero, the maximum possible difference in Type I and Type II errors is .001. As a practical matter, the difference is so small as to be inconsequential. That is, given the precision of the data, in terms of the number of significant digits to the right of the decimal, making a distinctions this small between the two types of error probabilities is likely not justified. Therefore, as a practical matter, the errors can still be viewed as balanced. It follows that implementing the ceiling will prevent the levels of significance of the parity tests from becoming outrageously small, with encountering little or no practical costs in terms of sacrificing the error balancing aspects of the procedure.

[15] Statistical Techniques For The Analysis And Comparison Of Performance Measurement Data. Submitted to Louisiana Public Service Commission (LPSC) Docket U-22252 Subdocket C

CLEC compared to the ILEC becomes larger, the absolute value of the modified z-statistic also becomes larger for the sample in the time period of interest. Thus, it is appropriate that the performance consequence should escalate based upon the calculated value of the modified z-statistic.

5. After a Failed Parity Test the Consequences Should Escalate and Vary Continuously with Severity of Failure

A parity failure is established for a submeasure by comparing the measured value of the modified z-statistic (z) to the balancing critical value (z*) appropriate for the submeasure's sample size during the given monthly period. Once a submeasure failure is obtained, the calculated remedy should be a continuous function of severity of the failure as measured by the magnitude of the modified z-statistic. In this way small changes in severity lead to small changes in consequences thus assuring that mathematically chaotic behavior is avoided at step thresholds. However, to incent the ILEC appropriately, the change in consequences should increase with each unit of severity. This form of consequences as a function of severity is most simply accomplished by the use of a quadratic function of the ratio of the measured modified z score to the balancing critical value (z/z*). Fixing the value of the quadratic or its slope at three points completely determines the function.

Table 1

Range of modified z-statistic value (z)	Performance Designation	Applicable Consequence ($)
greater than or equal z*	Compliant	0

[16] See Attachment D for a further discussion of this position.

less than z* to 5z*/3	Basic Failure	$a(z/z^*)^2 + b(z/z^*) + c$
less than 5z*/3 to 4z*	Intermediate Failure	
less than 4z*	Severe Failure	53,125

Table 1 shows the applicable consequences for each Tier I parity submeasure failure for each CLEC. In this table z* is the (negative) balancing critical value for the submeasure, and the coefficients of the smooth consequence function are:

a = 5625
b = -11250
c = 8125.

Note that the smooth consequences formula is an explicit function of the ratio of the modified z-statistic and the balancing critical value (z/z*). This means that the dollar amount does not depend on the number of observations but only on the degree of violation. If we had 100 times as many observations, with means and standard deviations staying the same, both z and z* will increase by a factor of 10 and the consequences will be unchanged. Note also that both basic and intermediate failures are defined and may occur in the smooth region of the formula. The plan retains these designations to allow for classification of performance for more general performance monitoring such as compliance testing, if needed.

Examples

Three hypothetical examples of consequence calculations are given in the matrix below.

Example	z*	z	Performance	Consequence

1	-2.00	-1.80	Compliant	$0
2	-2.50	-3.33	Basic Failure	$3,125
3	-3.00	-6.00	Intermediate Failure	$8,125
4	-2.50	-12.00	Severe Failure	$53,125

In example 1 the hypothetical balancing critical value for the submeasure is calculated to be −2.00 on the basis of sample size and equal type I and type II error probabilities. The observed value of the modified z-statistic, based on ILEC and CLEC performance for that submeasure, is −1.80. The ILEC is compliant for this submeasure and no consequences are due to this CLEC.

Example 2 shows a balancing critical value calculated to be −2.50. Furthermore in this example, the measured value of the modified z-statistic is −3.33. This is a Basic Failure and the consequence is calculated to be $3,125 by the formula in Table 1.

In example 3, although the hypothetical balancing critical value is −3.00, the measured value of the modified z-statistic is well below this at −6.00. According to the range of modified z-statistics in Table 1 this is an Intermediate Failure. The same smooth formula is used to calculate the remedy amount as $8,125.

The final example 4 shows a balancing critical value of -2.50, but a very poor measured value of the modified z-statistic of −12.00. According to Table 1 this is classified as a Severe Failure and generates a consequence of $53,125. This is the largest consequence for which the ILEC would be liable for this submeasure this month to this CLEC.

Tier I Business Rules for Benchmark Measurements

1. Use a "Bright Line" Test for Benchmark Measurements

A benchmark is set to define the level of performance that is judged essential to permit competition to develop on a going-forward basis. As such, the benchmark level is at the lower range of what a viable competitive support process should be capable of delivering on a routine basis. Indeed, to assume otherwise would imply that the benchmark would not be achieved on a routine basis. In all events, because even the most tightly controlled process will produce performance outside the expected range, some margin of error is typically provided for the incumbent. Thus, the limiting performance is expressed as "B% meet or exceed the benchmark" where "B%" is a proportion figure set less than 100% in order to account for random variation considerations. Accordingly, a performance failure should be declared if the calculated performance is not equal to the "B%" level. For example, if the calculated result for a month was 94.5% of all orders completed within 3 days but the benchmark was 95% within 3 days, then a performance failure occurred. No subsequent application of a statistical test is appropriate.

2. Apply an Adjustment for Small Data Sets When Necessary

Because some measurement results may be calculated using small data sets, some adjustment is warranted. This need arises because the benchmark proportion for a particular measure with few underlying data points may be practically impossible to attain unless the ILEC always performs perfectly. The metric discussed in the prior paragraph can be used to illustrate the point: if only ten orders were completed in the month, then compliance would occur only if all 10 orders were (correctly) completed within three days. One order taking longer than 3 days would mean that, at best, the performance result would be 90% within 3 days, i.e., a failing performance level.

This situation is addressed through application of the following table[17]:

Table 2

CLEC Data Set Size	Benchmark Percentage Adjustments for Small Data Sets (Applicable to Data Sets < 30)		
	85.0%	90.0%	95.0%
5	80.0%	80.0%	80.0%
6	83.3%	83.3%	83.3%
7	85.0%	85.7%	85.7%
8	75.0%	87.5%	87.5%
9	77.8%	88.9%	88.9%
10	80.0%	90.0%	90.0%
20	85.0%	90.0%	95.0%
30	83.3%	90.0%	93.3%

3. Increase Consequences for Increasingly Poor Performance

As with measurements that are judged against a parity standard, those compared to a benchmark standard should be subject to additional consequences as the performance becomes increasingly worse compared to the benchmark. The escalation is as follows (Note that "B" in Table 3, is the Benchmark Percentage as determined from Table 2):

Table 3

Range of Benchmark Result	Performance	Applicable Consequence ($)

[17] The table can be expanded to include all possible data set sizes from 1 upward.

(x)	**Designation**	
Meets or exceeds B%	Compliant	0
Meets or exceeds (1.5B-50)% but worse than B%	Basic Failure	$d[x/(100-B)]^2 + eB[x/(100-B)^2] + f[B/(100-B)]^2 + g$
Meets or exceeds (2.5B-150)% but worse than (1.5B-50)%	Intermediate Failure	
Worse than (2.5B-150)%	Severe Failure	53,125

In Table 3 the quantity x is the actually measured proportion and the coefficients are given by:

$d = 22500$

$e = -45000$

$f = 22500$

$g = 2500$

Example:

As an example of this consequence calculation, consider a benchmark with a proportion B=95%. Now if the measured performance is 93%, the first and second columns show that this is a Basic Failure. Plugging this 2% failure of the 95% benchmark proportion into the quadratic equation of the third column in the table gives a calculated consequence of $6,100 for this submeasure and CLEC.

Table 3 is applicable for any benchmark expressed as B% proportion better than L level, and all benchmarks may be easily expressed in this form.

Additional Tier I Business Rules Applicable to All Measurements

1. Increase Consequences for Chronic Performance Failures

Regardless of the type of measurement (parity or benchmark), if performance fails to achieve the Compliant level in consecutive reporting periods, then additional consequences should apply. The recommended treatment for chronic failures is to assess a chronic failure over-ride in the third consecutive month of non-compliant performance. When the chronic failure override applies, a consequence equal to a "Severe Failure" ($53,125 per chronic failure per month) should apply until such time as performance for the specific measurement result is again classified as Compliant.[18]

If performance fails to achieve the Compliant level after the chronic failure override applies for one month, the "Severe Failure" penalty shall be increased to $79,689.50 (1.5 · $53,125). If subsequent and concurrent failures occurs, the "Severe Failure" penalty shall increase by a factor of 1.5 until the Compliant level is attained. For example, if the Compliant level is not reached for the fifth concurrent month, the "Severe Failure " penalty shall be $119,531.25.

After two months of service at the Compliant level, the penalty shall return to its initial level with penalties based on the quadratic consequences formula. In the unlikely event that the chronic failure override is invoked for a second time, the "Severe Failure" penalty again shall increase by a factor of

[18] Alternatively, it is possible to institute consequences for repeated failures as early as the second consecutive month of failure. The amount of the consequence under such a structure would escalate more gradually. See Attachment A, Table A of MCI Worldcom and AT&T Joint Remedies Proposal Ex Parte filed in CC Docket 98-56, filed June 2, 1999.

1.5 (per month) until the Compliant level is attained. Once the compliant level is attained after the second chronic failure episode, the quadratic consequences formula shall be multiplied by a factor equal to the ratio of the maximum chronic failure penalty paid over the initial "Severe Failure" penalty(i.e. $53,125).[19] This scaling is necessary because repeated chronic failure overrides indicate that the penalty levels explicit in the consequences formula are too low. By factoring the consequences formula in this manner, the expected penalty of a failure of any degree is increased and the appropriate relationship between the penalty and severity level is maintained.

2. No Additional Protection of the ILEC is needed through Forgiveness Mechanisms or Mitigation Methods

Properly calibrated performance measures and balancing the probabilities of statistical errors eliminate any need for additional forms of protection for incumbents with respect to considerations of random variation.[20] Moreover, a procedural cap such as the one described below should allay any fears that additional protections are necessary for the ILEC.[21]

Tier II

[19] The consequences formula is linearly homogenous so that multiplying by a constant scales the minimum and maximum values of the initial formula by that constant. For example, if the minimum penalty is $2,500 and the maximum $53,125, then multiplying by a factor of 1.5 increases the minimum penalty to $3,750.00, the maximum penalty to $79,687.50, and the quadratic formulation will produce a continuous penalty amount between those two values based on the ratio z/z*.

[20] See Attachment E for further discussion of random variation and the inappropriateness of providing further mitigation if Type I and Type II error is balanced as recommended in this proposal.

[21] Because the rationale for providing consequence offsets is the possibility of random variation, there is no justification for applying offsets to measurements that are monitored through the use of benchmarks. As explained above, random variability impacts are fully

Tier II consequences are intended to enhance the ILEC's incentives to provide performance that complies with its statutory obligations. Tier I consequences only compensate individual CLECs who actually receive discriminatory treatment from the ILEC. Tier II consequences are designed to counterbalance the ILEC's incentive to damage not just individual firms but the competitive marketplace itself. Thus, the two types of consequences are complementary, and both are necessary to achieve the intended results.

The applicability of Tier II consequences should be determined using the aggregate data for all CLECs within a particular submeasurement result and disaggregation.[22] Except as noted below, identical business rules and measurements should be utilized as for Tier I. Thus, virtually the same data and computational processes can be utilized for both tiers. The differences are highlighted below and are due largely to a reduction of the consequence threshold below the balancing critical value. The smaller threshold is recommended because higher consequences are proposed, so the confidence in the decision to apply a consequence should be greater.

Because Tier II consequences reflect harm to the public interest in a competitive marketplace, consequences under Tier II, unlike Tier I payments, should be paid to a public fund identified by the Commission and may be used for competitively neutral public purposes.[23]

Tier II Business Rules for Parity Measurements.

---

cared for in the structure of the benchmark standard, by permitting in advance a percentage of performance "misses."

[22] Each occurrence counts equally in this calculation. Thus, the individual results for individual CLECs are not averaged together; rather the performance for all CLECs is pooled for each submeasurement result. Thus the pooled data analysis effectively creates a "super CLEC" for the purposes of determining Tier II consequences.

[23] Thus, under Tier II, individual CLECs are not compensated.

The same business rules apply under Tier II to the aggregate (or pooled) data of the individual CLECs as are employed for the individual CLEC data under Tier I, except a smaller consequence threshold is used.[24] As a result, the applicable consequence table (Table 1 above) is modified as follows:

Table 4

Range of modified z-statistic value (z)	Performance Designation	Applicable Consequence ($)
greater than or equal $5z^*/3$	Indeterminate	0
less than $5z^*/3$ to $4z^*$	Market Impacting	$n\,[a(z/z^*)^2 + b(z/z^*) + c]$
less than $4z^*$	Market Constraining	n53,125

Here z^* is the balancing critical value for the given submeasure aggregated over all the CLECs, and the coefficients of the smooth consequence function are again:

$a = 5625$

$b = -11250$

$c = 8125.$

The quantity n is the market penetration factor explained below.

[24] Alternative methodology exists for determining Tier II consequences. See, for example, the June 2, 1999 Joint AT&T and MCI ex parte filing made with the FCC in CC Docket 98-56.

Tier II Business Rules for Benchmark Measurements

The same business rules apply under Tier II to the aggregate (or pooled) data of the individual CLECs as are employed for the individual CLEC data under Tier I, except that consequences do not apply until the pooled CLEC performance results degrades to a point that is equivalent to an intermediate failure designation at the Tier I level. As with parity measures, the applicable consequences are adjusted to reflect the broader consequences of poor performance for the entire CLEC industry and the concomitant effects on the market and consumers.

Table 5

Range of Benchmark Result (x)	Failure Designation	Applicable Consequence ($)
Meets or exceeds (1.5B-50)%	Indeterminate	0
Meets or exceeds (2.5B-150)% but worse than (1.5B-50)%	Market Impacting	n {d[x/(100-B)]^2 + eB[x/(100-B)^2] + f[B/(100-B)]^2 + g}
Worse than (2.5B-150)%	Market Constraining	n53,125

For Table 5, x is the actually measured proportion and the coefficients are again given by:

d = 22500

e = -45000

f = 22500

g = 2500

The quantity n is the market penetration factor explained below.

Establishing the Value of "n" for Tier II

For both Tier II tables (Tables 4 and 5), the value for "n" should be determined based upon the most recent data for the state and company under consideration (in this caseFlorida) relating to resold lines (Table 3.1) and UNE loops (Table 3.3) as reported in the most recent Report of Local Competition published by the FCC.[25] In effect, "n" is a multiplier for the Tier II consequence amount that takes into account, in general terms, the extent of competitive penetration within the state.[26]

Table 6

Lines provided to CLECs/Total ILEC and CLEC Lines	Value of "n"
more than 50%	0
more than 40% to less than or equal 50%	1
more than 30% to less than or equal 40%	2
more than 20% to less than or equal 30%	4
more than 10% to less than or equal 20%	6

[25] If a company is not explicitly identified, then the aggregate result for the state would be utilized

[26] The calculation for a particular ILEC and state would be based on the most current data reported to the FCC and be as follows: (resold lines + UNE loops)/(total switched lines).

more than 5% to less than or equal 10%	8
0% to less than or equal 5%	10

Thus, as competition becomes established, the size of the applicable Tier II consequence is reduced to zero if the ILEC no longer provides a majority of the local lines to the CLECs in its serving area.

Other Considerations

1. **Procedural Caps May Be Useful If Properly Implemented**

In the course of early state consideration of consequence plans, regulators and incumbents expressed concern regarding the possible size of payments that an incumbent might be required to pay. In response, proposals were made to cap incumbents' potential liability. As a threshold matter, it should be noted that this concern reflects a tacit acknowledgement that the performance delivered by the incumbents has to date been largely non-complaint. Moreover, to the extent that any cap is considered at all, the very important difference between absolute and procedural caps must be recognized. As shown below, if the Commission establishes any caps at all, they should be purely procedural and not place an absolute limit on the potential consequence payments due from the ILEC.[27]

The difference between procedural and absolute caps is significant. Absolute caps should be avoided entirely. First, such caps provide an ILEC with the means to evaluate the cost of market share retention through delivery of non-compliant performance. Second, absolute caps send the

[27] In this regard, it should be noted that the main purpose of any system of incentives is to have an ILEC accept its legal responsibility to perform at appropriate levels and not pay any consequences at all.

signal that once the ILEC's performance deteriorates to a particular level (i.e., reaching the absolute cap) then further deterioration is irrelevant.[28]

Procedural caps, on the other hand, establish a preset level at which the ILEC could seek regulatory review of the consequences that are due; however, the cap would not automatically absolve an ILEC of liability for a consequence. Procedural caps, therefore, avoid both of the problems of absolute caps. They do not provide ILECs with the opportunity to evaluate the "cost" of retaining share through non-compliance. Likewise, they do not absolve an ILEC from consequences for unchecked performance deterioration.

To the extent a procedural cap is employed, it should be tailored to achieve the following:

> (1) A meaningful level of consequences must be available before the procedural cap applies;
> (2) The procedural cap should apply on a rolling twelve-month period and not to individual months;
> (3) The procedural cap should not apply to Tier I consequences for the CLECs but only Tier II consequences.[29] No other caps should be applicable.
> (4) To the extent that a procedural cap is exceeded, the ILEC must pay out consequences up to the procedural cap and put the amount in

[28] Similarly, the use of weightings for individual performance measurements to determine the amount of consequences should also be avoided. Any weighting process is inherently subjective and thus arbitrary. Moreover, use of weightings may inappropriately influence the market entry mode selected by a particular CLEC. It is far superior to permit the market to determine which measures are most important by seeing what functions customers need from CLECs, and that CLECs in turn need from the ILEC.

excess of the cap in an escrow account that earns a minimum interest rate as approved by the Commission;

(5) The Commission shall decide whether and to what extent the amount in excess of the procedural cap should be paid out. The ILEC should pay out any amount in excess of the cap, including accrued interest, according to Commission order.

The level of the procedural cap must be set high enough that meaningful incentives are immediately payable without intervention of the Commission. To permit otherwise would effectively prevent the performance consequences from being self-enforcing. It is reasonable to expect that any procedural cap should be proportionate to the size of the local market at issue. It is therefore recommended that, if a procedural cap is adopted, that it be determined from the estimated dollar amount that the ILEC stands to retain in monopoly based revenues.[3031]

2. <u>Other Provisions Protect ILECs From The Impact Of Extraordinary Events</u>

The cut of a single cable may result in higher trouble rates and longer mean times to repair over a short period of time. This is referred to as clustering. While clustering may in fact occur, there is no particular reason to believe that any such events would result in disproportionate impacts on the ILECD or even the CLECs. Furthermore, there may be other events demonstrably

[29] As noted above, Tier I consequences principally act as a form of liquidated damages. Thus, there is no justification for capping such consequences whether for an individual CLEC or for the CLEC industry as a whole.

[30] See Affidavit of R. Glenn Hubbard and William H. Lehr on behalf of AT&T Corp. AT&T Exhibit _ before the Federal Communications Commission, Washington, D.C. 20544, in the matter of application by New York Telephone Company (d/b/a Bell Atlantic-New York). CC Docket no. 99-295.

beyond the control of the ILEC that may affect its service quality differently from the CLECs'. This condition does not argue that automatic exclusion should be provided for an otherwise applicable consequence. Nevertheless, the ILEC should not be denied protection from extraordinary impacts not anticipated in the construction of the consequence plan[32]. As a result, if such events occur, the ILECshould be permitted to pursue relief according to the following:

(1) The ILEC should notify the Commission and any potentially affected CLEC(s), using written and verifiable means of notice, of the intent to pursue an exception. Such notification must be provided before the applicable consequence is payable; otherwise the ILEC waives its rights.

(2) All consequences not at issue under the exception petition must be immediately payable as provided for elsewhere in the plan. Those that are subject of the potential exemption shall be paid into an interest bearing escrow account no later than the due date applicable to the consequences that are at issue.

(3) No later than 15 calendar days following the due date of the consequences for which an exemption is sought, the incumbent shall submit to the Commission and all other affected parties all factual evidence

---

[31] SBC in Texas has agreed to a $120M annual limit for consequences where 9M lines are in service

[32] Root cause analysis should not defer payments of consequences. ILECs must be liable to pay any consequences for poor performance. Completion of root cause analysis must not be a prerequisite for the delivery of payments to either the CLEC(s) or to the designated Tier II fund. Root cause analyses tend to be time consuming to conduct. While root cause analysis is desirable for long range performance improvement purposes, it is antithetical to self-enforcing consequences. Finally, the provisions set forth in the immediately preceding section provide a procedural mechanism available to ILECs should after-the-fact root cause analysis indicate that a consequence was misapplied from the ILEC's perspective.

supporting the exemption. To the extent the ILEC seeks proprietary protection of the information submitted, it shall employ a standard nondisclosure form, approved by the Commission, before the plan is put into operation. The ILEC may not rely upon the lack of the proprietary form as a basis to delay the submission to the Commission, nor may the incumbent delay access to information by any CLEC that agrees to sign the standard nondisclosure form.

(4) By the later of 30 calendar days following notice by the incumbent or 15 calendar days following the ILEC's compliance with (3) above, interested CLECs shall file comments regarding the requested exemption. By mutual agreement, this period may be extended up to 15 calendar days.

(5) Following closure of the comment period provided in (4), if the ILEC and CLEC(s) have not reached a mutually agreeable settlement, the Commission shall either

(a) render a decision regarding the requested exemption, or

(b) seek further comment. The Commission shall render its decision regarding the exemption, which shall be binding on all parties, within 90 calendar days of the payment due date of the consequences at issue.

(6) Payout of the consequences shall be according to Commission direction and liquidate the entire escrow account, including accrued interest. In addition, the ILEC should be responsible for reimbursing reasonably incurred legal fees of the CLECs. Such amounts should be reimbursed in the following proportion:

[1-(amount returned to the incumbent)]/total escrow balance at liquidation.

As discussed in Attachment F, other steps may be taken to address potential measurement correlation issues once actual data has been gathered under the performance measurement system.

3. Additional Consequences Enforce the Operation of the Plan

Additional consequences should be applicable for other ILEC failures related to performance reporting. At a minimum, consequences for the following areas of non-compliance are appropriate:

Late performance reports - If performance data and associated reports are not available to the CLECs by the due day, the ILEC should be liable for payments of $5,000 to a state fund for every day past the due date for delivery of the reports and data. The ILEC's liability should be determined based on the latest report delivered to a CLEC.

Incomplete or revised reports - If performance data and reports are incomplete, or if previously reported data are revised, then the ILEC should be liable for payments of $1,000 to a state fund for every day past the due date for delivery of the original reports.

Inability to access detailed data - If a CLEC cannot access its detailed data underlying the ILEC's performance reports due to failures under the control of the ILEC, then the ILEC should pay the affected CLEC $1000 per day (or portion thereof) until such data are made available.

Interest on late consequence payments - If the ILEC fails to remit a consequence payment by the 15th business day following the due date of the data and the reports upon which the consequences are based, then it should be liable for accrued interest for every day that the payment is late. A per diem interest rate that is equivalent to the ILEC's rate of return for its regulated services for the most recent reporting year should apply.

Attachment A

Sufficient Disaggregation Is Essential to Permit Detection of Discrimination

A meaningful system of performance consequences cannot operate without a high-quality system of performance measurements. This requires not only a robust system of performance measurements that monitors all key aspects of market entry and ILEC support but also that the results derived from such measurements are sufficiently discrete to permit meaningful comparisons.[33]

Sufficient disaggregation is absolutely essential for accurate comparison of results to expected performance. This is true regardless of whether parity or a benchmark serves as the performance standard. Inadequate disaggregation of results means that not all key factors driving differences in performance results have been identified, which in turn interjects needless variability into the computed results. Such an outcome has two adverse effects. First, the ability to detect real differences is reduced for parity measures, because the modified z-statistic employs only the incumbent's variance in the denominator, which will increase with inappropriate averaging of dissimilar results (thus causing the calculated z-statistic to be smaller). Second, benchmark standards may be more permissive, both in terms of the absolute standard and the percentage "miss" accepted (to the extent it is factually supported at all), if the factual data underlying them are averages of widely divergent processes. Accordingly, inadequately disaggregated data impose very lenient targets that result in a very low probability that performance requirements will be missed.

[33] Although some incumbents have raised vague concerns that sufficient disaggregation of results may over-burden regulators, those concerns are unfounded for two reasons. First, careful advance specification of disaggregation requirements will reduce, rather than increase, regulatory burden and permit superior quality decision making. Second, if fewer performance results are desired, statistical procedures for re-aggregating disaggregated results provide a superior approach to reliance upon overly aggregated measurement results.

Only incumbents, such as BellSouth, have access to the highly detailed information regarding their retail performance necessary to determine the level of disaggregation that is required to permit apples-to-apples comparisons. Moreover, there are analytical procedures that allow factual conclusions to be made regarding how much disaggregation is "enough."[34] Indeed, in the limited instances where CLECs have been provided access to ILEC data and at least limited public disclosure of analysis was permitted, the facts showed both that ILECs have very detailed data and that very disaggregated results comparisons are necessary to avoid bias.[35] Establishing the appropriate level of disaggregation is not a "once-and-done" undertaking. Provision can be made to review, perhaps annually, the appropriateness of the disaggregation contained in the ILEC's performance measurement system. In this review process, an ILEC may demonstrate, through data it has collected pursuant to its performance measurement system, that the existing level of disaggregation is not providing any additional insight to an assessment of its performance quality and nondiscrimination. In that same review process, individual CLECs should also be permitted to request additional disaggregation.[36] The party requesting a change should have the burden of showing why the proposed change is appropriate provided that all parties have equal access to detailed data necessary to support the proposal.

There should not be any presumption that additional disaggregation creates a burden, for either the ILEC or this Commission. For all incumbents in

[34] For example, regression procedures may provide a workable methodology for establishing the extent of disaggregation required to make accurate comparisons.
[35] See AT&T Ex Parte filed July 20, 1999 in CC Docket 98-56.
[36] In such cases, the requesting CLEC should be required to make its request for further disaggregation to the incumbent LEC at least three months before initiation of the review process.

general, additional disaggregation (once correct implementation is validated) simply involves repetitive computation – a task readily and quickly accomplished by today's computers. Such a small and largely one-time effort is a small price to pay for the vastly improved capability to protect the prospects for competition in Florida.

Attachment B

Local Competition Users Group

Statistical Tests for Local Service Parity

February 6, 1998
Membership: AT&T, Sprint, MCI, LCI, WorldCom

Version 1.0

Executive Summary

The Local Competition Users Group has drafted 27 Service Quality Measurements (SQMs) that will be used to measure parity of service provided by incumbent local exchange carriers (ILECs) to competitive local exchange carriers (CLECs). This set of measures includes means, proportions, and rates of various indicators of service quality. This document proposes statistical tests that are appropriate for determining if parity is being provided with respect to these measurements.

Each month, a specified report of the 27 SQMs will be provided by the ILEC, broken down by the requested reporting dimensions. The SQMs are to be systematically developed and provided by the ILECs as specified. Test parameters will be calculated so that the overall probability of declaring the ILEC to be out of parity purely by chance is very small. For each SQM and reporting dimension reported, the difference between the ILEC and CLEC results is converted to a *z*-value. Non-parity is determined if a *z*-value exceeds a selected critical value.

Introduction

Purpose

The Local Competition Users Group (LCUG) is a cooperative effort of AT&T, MCI, Sprint, LCI and WorldCom for establishing standards for the entry of new companies (competitive local exchange carriers, or CLECs) into the local telecommunications market. A key initiative of the LCUG is to establish measures of parity for services provided by incumbent local exchange carriers (ILECs). In short, parity means that the support ILECs provide on behalf of the CLECs is no lesser in quality than the service provided by the ILECs to their own customers.

The LCUG has drafted a document listing service quality measurements (SQMs) that must be reported by the ILECs to insure that CLECs are given parity of support. The SQM document has been submitted to the FCC and made available to PUCs in all 50 states and is pending approval by many of these regulatory agencies. This document has been drafted to describe statistical methodology for determining if parity exists based on the measurements defined in the SQM document.

Service Quality Measurements

The LCUG has identified 27 service quality measurements for testing parity of service. These are:

Category	ID	Description
Pre-Ordering	PO-1	Average Response Interval for Pre-Ordering Information
Ordering and Provisioning	OP-1	Average Completion Interval
	OP-2	Percent Orders Completed on Time
	OP-3	Percent Order Accuracy
	OP-4	Mean Reject Interval
	OP-5	Mean FOC Interval
	OP-6	Mean Jeopardy Interval
	OP-7	Mean Completion Interval
	OP-8	Percent Jeopardies Returned
	OP-9	Mean Held Order Interval
	OP-10	Percent Orders Held >= 90 Days
	OP-11	Percent Orders Held >= 15 Days
Maintenance and Repair	MR-1	Mean Time to Restore
	MR-2	Repeat Trouble Rate
	MR-3	Trouble Rate

	MR-4	Percentage of Customer Troubles Resolved Within Estimate
General	GE-1	Percent System Availability
	GE-2	Mean Time to Answer Calls
	GE-3	Call Abandonment Rate
Billing	BI-1	Mean Time to Provide Recorded Usage Records
	BI-2	Mean Time to Deliver Invoices
	BI-3	Percent Invoice Accuracy
	BI-4	Percent Usage Accuracy
Operator Services and Directory Assistance	OSDA-1	Mean Time to Answer
Network Performance	NP-1	Network Performance Parity
Interconnect / Unbundled Elements and Combos	IUE-1	Function Availability
	IUE-2	Timeliness of Element Performance

The Service Quality Measurements document describes the importance of each measure as an indicator of service parity. The SQM document also describes reporting dimensions that will be used to break each measure out by like factors (*e.g.*, major service group).

Why We Need to Use Statistical Tests

The Telecommunications Act of 1996 requires that ILECs provide nondiscriminatory support regardless of whether the CLEC elects to employ interconnection, services resale, or unbundled network elements as the market entry method. It is essential that CLECs and regulators be able to determine whether ILECs are meeting these parity and nondiscriminatory obligations. In order to make such a determination, the ILEC's performance for itself must be compared to the ILEC's performance in support of CLEC operations; and the results of this comparison must demonstrate that the CLEC receives no less than equal treatment compared to that the ILEC provides to its own operations. Where a direct comparison to analogous ILEC performance is not possible, the comparative standard is the level of performance that offers an efficient CLEC a meaningful opportunity to compete.

When making the comparison of ILEC results to CLEC results, it is necessary to employ comparative procedures that are based upon generally accepted statistical procedures. It is important to use statistical procedures because all of the ILEC-CLEC processes that will be measured are processes that contain some degree of randomness. Statistical procedures recognize that there is measurement variability, and assist in translating results data into useful decision-making information. A statistical approach allows for measurement variability while controlling the risk of drawing an inappropriate conclusion (*i.e,* a "type 1" or "type 2" error, discussed in the next section).

Basic Concepts and Terms

Populations and Samples

Statistical procedures will permit a determination whether the support that the ILECs provide to CLECs is indistinguishable from the support provided by the ILECs to their own customers. In statistical terms, we will determine whether two "samples", the ILEC sample and the CLEC sample, come from the same "population" of measurements.

The procedures described in this paper are based on the following assumption: *When parity is provided, the ILEC data and CLEC data can both be regarded as samples from a common population of possible outcomes.* In other words, if parity exists, the measured results for a CLEC should not be distinguishable from the measured results for the ILEC, once random variability is taken into account. Figure 1 illustrates this concept. On the right side of the figure are histograms of two samples. In this illustration, the ILEC sample contains 200 observations (data values) and the CLEC sample contains 50. Note that the two histograms are not exactly alike. This is due to sampling variation. The assumption that parity exists implies that both samples were drawn from the same population of values. If it were possible to observe this population completely, the population histogram might appear as shown on the left of the Figure. If the samples were indeed taken from this population, histograms drawn for larger and larger samples would look more and more like the population histogram. Figure 1 shows that even when parity is being provided, there will be differences between the samples due to sampling variability. Statistical tests quantify the differences between the two samples and make proper allowance for sampling variability. They assess the chance that the differences that are observed are due simply to sampling variability, if parity is being provided.

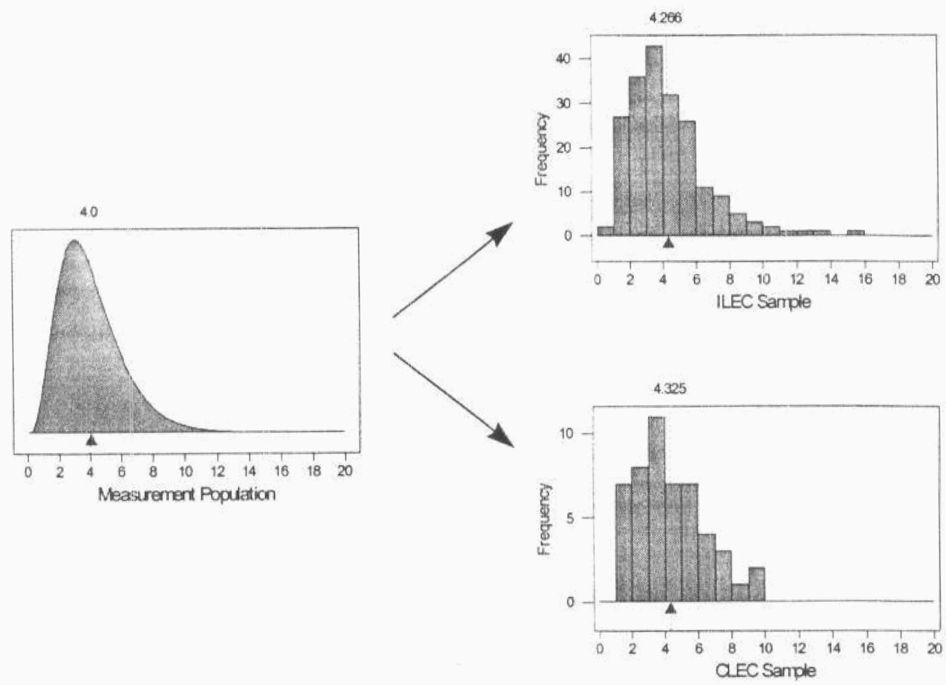

Figure 1.

Measures of Central Tendency and Spread

Often, distributions are summarized using "statistics." For the purpose of this paper, a "statistic" is simply a calculation performed on a sample set of data. Two common types of statistics are known as measures of "central tendency" and "spread."

A measure of central tendency is a summary calculation that describes the middle of the distribution in some way. The most common measure of central tendency is called the "mean" or "average" of the distribution. The mean of a sample is simply the sum of the data values divided by the sample size (number of observations). Algebraically, this calculation is expressed as

$$\bar{x} = \frac{\sum x}{n},$$

where x denotes a value in the sample and n denotes the sample size. The mean describes the center of the distribution in the following way: *If the histogram for a sample were a set of weights stacked on top of a flat board placed on top of a fulcrum (a "see-saw"), the mean would be the position along the board at which the board would balance.* (See Figure 1.) The mean in Figure 1 is indicated by the small triangle at approximately the value "4" on the horizontal axis.

A measure of spread is a summary calculation that describes the amount of variation in a sample. A common measure of spread is a called the "standard deviation" of the sample. The standard deviation is the typical size of a deviation of the observations in the sample from their mean value. The standard deviation is calculated by subtracting the mean value from each observation in the sample, squaring the resulting differences (so that negative and positive differences don't offset), summing the squared differences, dividing the sum by one less than the sample size, then taking the square root of the result. Algebraically, this calculation is expressed as

$$\sigma = \sqrt{\frac{\Sigma(x - \bar{x})^2}{n - 1}}.$$

While the notion of mean and standard deviation exists for populations as well as samples, the mathematical definition for the mean and standard deviation for populations is beyond the scope of this paper. However, their interpretation is generally the same as for samples. In fact, for very large samples, the sample mean and sample standard deviation will be very close to the mean and standard deviation of the population from which the sample was taken.

Sampling Distribution of the Sample Mean

In Figure 1 we showed the positions of the means of the population and the two samples with triangular symbols beneath the distributions. If we sample over successive months, we will get new ILEC samples and new CLEC samples each and every month. These samples will not be exactly like the one for the first month; each will be influenced by sampling variability in a different way. In Figure 2, we show how sets of 100 successive ILEC means and 100 successive CLEC means might appear. The ILEC means can be thought of as being drawn from a population of sample means; this population is called the "sampling distribution" of these ILEC means. This sampling distribution is completely determined by the basic population of measurements that we start with, and the number of observations in each sample. The sampling distribution has the same mean as the population.

Figure 2 illustrates two important statistical concepts:

1. The histogram of successive sample means resembles a bell-shaped curve known as the Normal Distribution. This is true even though the individual observations came from a skewed distribution.
2. The standard deviation of the distribution of sample means is much smaller than the standard deviation of the observations themselves. In fact, statistical theory establishes the fact that the standard deviation on

the population of means is smaller by a factor $\sqrt{n}$, where n is the sample size. This effect can be seen in our example: the distribution of the CLEC means is twice as broad as the distribution of the ILEC means, since the ILEC sample size (200) is four times as large as the CLEC sample size (50).

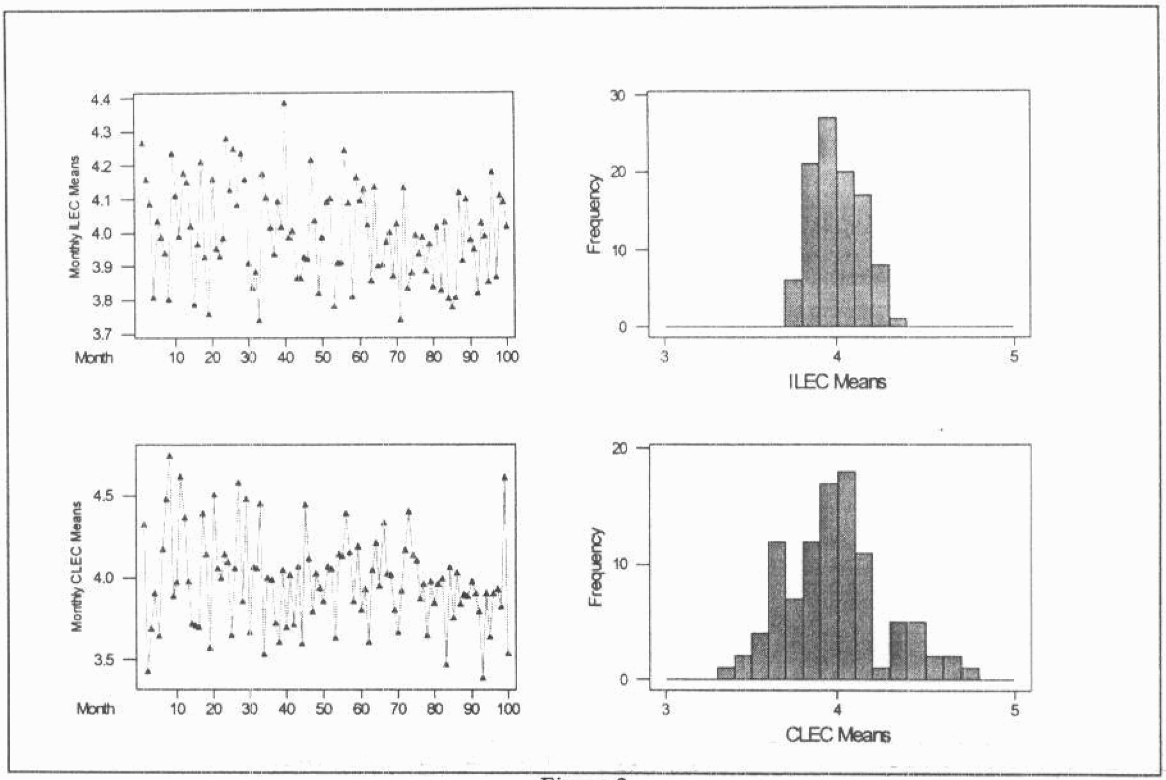

Figure 2.

It is common to call the standard deviation of the sampling distribution of a statistic the "standard error" for the statistic. We shall adopt this convention to avoid confusion between the standard deviation of the individual observations and the standard deviation (standard error) of the statistic. The latter is generally much smaller than the former. In the case of sample means, the standard error of the mean is smaller than the standard deviation of the individual observations by a factor of $\sqrt{n}$.

The Z-test

Our objective is to compare the mean of a sample of ILEC measurements with the mean of a sample of CLEC measurements. Suppose both samples were drawn from the same population; then the difference between these two sample means (*i.e.*, $DIFF = \bar{x}_{CLEC} - \bar{x}_{ILEC}$) will have a sampling distribution which will

(i) have a mean of zero; and

(ii) have a standard error that depends on the population standard deviation and the sizes of the two samples.

Statisticians utilize an index for comparing measurement results for different samples. The index employed is a ratio of the difference in the two sample means (being compared) and the standard deviation estimated for the overall population. This ratio is known as a z-score. The z-score compares the two samples on a standard scale, making proper allowance for the sample sizes.

The computation of the difference in the two sample means is straightforward.

$$DIFF = \bar{x}_{\text{CLEC}} - \bar{x}_{\text{ILEC}}$$

The standard deviation is less intuitive. Nevertheless, statistical theory establishes the fact that

$$\sigma^2_{\text{DIFF}} = \frac{\sigma^2}{n_{\text{CLEC}}} + \frac{\sigma^2}{n_{\text{ILEC}}},$$

where is the standard deviation of the population from which both samples are drawn. That is, the squared standard error of the difference is the sum of the squared standard errors of the two means being compared.[37]

We do not know the true value of the population because the population cannot be fully observed. However, we can estimate given the standard deviation of the ILEC sample ($_{\text{ILEC}}$).[38] Hence, we may estimate the standard error of the difference with

$$\sigma_{\text{DIFF}} = \sqrt{\frac{\sigma^2_{\text{ILEC}}}{n_{\text{CLEC}}} + \frac{\sigma^2_{\text{ILEC}}}{n_{\text{ILEC}}}} = \sqrt{\sigma^2_{\text{ILEC}}\left[\frac{1}{n_{\text{CLEC}}} + \frac{1}{n_{\text{ILEC}}}\right]}$$

If we then divide the difference between the two sample means by this estimate of the standard deviation of this difference, we get what is called a "z-score".

$$z = \frac{DIFF}{\sigma_{\text{DIFF}}}$$

[37] Winkler and Hays, *Probability, Inference, and Decision*. (Holt, Rinehart and Winston: New York), p. 370.

[38] Winkler and Hays, *Probability, Inference, and Decision*. (Holt, Rinehart and Winston: New York), p. 338.

Because we assumed that both samples were in fact drawn from the same population, this *z*-score has a sampling distribution that is very nearly Standard Normal, *i.e.*, having a mean of zero and a standard error of one. Thus, the *z*-score will lie between ± 1 in about 68% of cases, will lie between ± 2 in about 95% of cases, and will lie between ± 3 in about 99.7% of cases, always assuming that both samples come from the same population. Therefore, one possible procedure for checking whether both samples come from the same population is to compare the *z*-score with some cut-off value, perhaps +3. For comparisons where the values of z exceed the cutoff value, you reject the assumption of parity as not proven by the measured results. This is an example of a statistical test procedure. It is a formal rule of procedure, where we start with raw data (here two samples, ILEC measurements and CLEC measurements), and arrive at a decision, either "conformity" or" violation".

Type 1 Errors and Type 2 Errors

Each statistical test has two important properties. The first is the probability that the test will determine that a problem exists when in fact there is none. Such a mistaken conclusion is called a type one error. In the case of testing for parity, a type one error is the mistake of charging the ILEC with a parity violation when they may not be acting in a discriminatory manner. The second property is the probability that the test procedure will not identify a parity violation when one does exist. The mistake of not identifying parity violation when the ILEC is providing discriminatory service is called a type two error. A balanced test is, therefore, required.

From the ILEC perspective, the statistical test procedure will be unacceptable if it has a high probability of type one errors. From the CLEC perspective, the test procedure will be unacceptable if it has a high probability of type two errors.

Very many test procedures are available, all having the same probability of type one error. However the probability of a type two error depends on the particular kind of violation that occurs. For small departures from parity, the probability of detecting the violation will be small. However, different test procedures will have different type two error probabilities. Some test procedures will have small type two error when the CLEC mean is larger than the ILEC mean, even if the CLEC standard deviation is the same as the ILEC standard deviation, while other procedures will be sensitive to differences in standard deviation, even if the means are equal. Our proposals below are designed to have small type two error when the CLEC mean exceeds the ILEC mean, whether or not the two variances are equal.

Tests of Proportions and Rates

When our measurements are proportions (*e.g.* percent orders completed on time) rather than measurements on a scale, there are some simplifications. We can think of the "population" as being analogous to an urn filled with balls, each labeled either 0(failure) or 1(success). In this population, the fraction of 1's is some "population proportion". Making an observation corresponds to drawing a single ball from this urn. Each month, the ILEC makes some number of observations, and reports the ratio of failures or successes to the total number of observations; the ILEC does the same does the same for the CLEC. The situation is very similar to that discussed above; however, rather than a wide range of possible result values, we simply have 0's (failures) and 1's (successes). The "sample mean" becomes the "observed proportion", and this will have a sampling distribution just as before. The novelty of the situation is that now the population standard deviation is a known function of the population proportion[39]; if the population proportion is p, the population standard deviation is $\sqrt{p(1-p)}$, with similar simplifications in all the other formulas.

There is a similar simplification when the observations are of rates, *e.g.*, number of troubles per 100 lines. The formulas appear below.

Proposed Test Procedures

Applying the Appropriate Test

Three z-tests will be described in this section: the "Test for Parity in Means", the "Test for Parity in Rates", and the "Test for Parity in Proportions". For each LCUG Service Quality Measurement (SQM), one or more of these parity tests will apply. The following chart is a guide that matches each SQM with the appropriate test.

Measurement (Corresponding LCUG Number)	*Test*
Preordering Response Interval (PO-1)	**Mean**
Avg. Order Completion Interval (OP-1)	**Mean**
% Orders Completed On Time (OP-2)	**Proportion**
% Order (Provisioning) Accuracy (OP-3)	**Proportion**
Order Reject Interval (OP-4)	**Mean**
Firm Order Confirmation Interval (OP-5)	**Mean**
Mean Jeopardy Interval (OP-6)	**Mean**
Completion Notice Interval (OP-7)	**Mean**
Percent Jeopardies Returned (OP-8)	**Proportion**
Held Order Interval (OP-9)	**Mean**

[39] Winkler and Hays, *Probability, Inference, and Decision*. (Holt, Rinehart and Winston: New York), p. 212.

% Orders Held ≥ 90 Days (OP-10)	**Proportion**
% Orders Held ≥ 15 Days (OP-11)	**Proportion**
Time To Restore (MR-1)	**Mean**
Repeat Trouble Rate (MR-2)	**Proportion**
Frequency of Troubles (MR-3)	**Rate**
Estimated Time To Restore (MR-4)	**Proportion**
System Availability (GE-1)	**Proportion**
Center Speed of Answer (GE-2)	**Mean**
Call Abandonment Rate (GE-3)	**Proportion**
Mean Time to Deliver Usage Records (BI-1)	**Mean**
Mean Time to Deliver Invoices (BI-2)	**Mean**
Percent Invoice Accuracy (BI-3)	**Proportion**
Percent Usage Accuracy (BI-4)	**Proportion**
OS/DA Speed of Answer (OS/DA-1)	**Mean**
Network Performance (NP-1)	**Mean, Proportion**
Availability of Network Elements (IUE-1)	**Mean, Proportion**
Performance of Network Elements (IUE-2)	**Mean, Proportion**

Test for Parity in Means

Several of the measurements in the LCUG SQM document are averages (*i.e.*, means) of certain process results. The statistical procedure for testing for parity in ILEC and CLEC means is described below:

1. Calculate for each sample the number of measurements (n_{ILEC} and n_{CLEC}), the sample means ($\bar{x}_{\text{ILEC}}$ and $\bar{x}_{\text{CLEC}}$), and the sample standard deviations (ILEC and CLEC).

2. Calculate the difference between the two sample means; if *larger* CLEC mean indicates possible violation of parity, use *DIFF* = $\bar{x}_{\text{CLEC}} - \bar{x}_{\text{ILEC}}$, otherwise reverse the order of the CLEC mean and the ILEC mean.

3. To determine a suitable scale on which to measure this difference, we use an estimate of the population variance based on the ILEC sample, adjusted for the sized of the two samples: this gives the standard error of the difference between the means as

$$\sigma_{\text{DIFF}} = \sqrt{\sigma^2_{\text{ILEC}}\left[\frac{1}{n_{\text{CLEC}}} + \frac{1}{n_{\text{ILEC}}}\right]}$$

4. Compute the test statistic

$$z = \frac{DIFF}{\sigma_{\text{DIFF}}}$$

5. Determine a critical value *c* so that the type one error is suitably small.

6. Declare the means to be in violation of parity if $z > c$.

Example:

c:	3.58	Critical value for the test

ILEC			CLEC			Test	
n	mean	variance	n	mean	variance	z	Violation
250	4.038	1.9547	50	5.154	23.2035	5.15	YES!

Test for Parity in Proportions

Several of the measurements in the LCUG SQM document are proportions derived from certain counts. The statistical procedure for testing for parity in ILEC and CLEC proportions is described below. It is the same as that for means, except that we do not need to estimate the ILEC variance separately.

1. Calculate for each sample sample sizes (n_{ILEC} and n_{CLEC}), and the sample proportions (p_{ILEC} and p_{CLEC}).

2. Calculate the difference between the two sample means; if *larger* CLEC proportion indicates worse performance, use $DIFF = p_{\text{CLEC}} - p_{\text{ILEC}}$, otherwise reverse the order of the ILEC and CLEC proportions.

3. Calculate an estimate of the *standard error for the difference* in the two proportions according to the formula

$$\sigma_{\text{DIFF}} = \sqrt{p_{\text{ILEC}}(1 - p_{\text{ILEC}})\left[\frac{1}{n_{\text{CLEC}}} + \frac{1}{n_{\text{ILEC}}}\right]}$$

4. Hence compute the test statistic

$$z = \frac{DIFF}{\sigma_{\text{DIFF}}}$$

5. Determine a critical value c so that the type one error is suitably small.

6. Declare the means to be in violation of parity if $z > c$.

Example:

c:	3.58	Critical value for the test

ILEC			CLEC			Test	
num	den	p	num	den	p	z	Violation
5	250	2.00%	7	40	17.50%	6.50	YES!

Test for Parity in Rates

A rate is a ratio of two counts, *num/denom*. An example of this is the trouble rate experience for POTS. The procedure for analyzing measurements results that are rates is very similar to that for proportions.

1. Calculate the numerator and the denominator counts for both ILEC and CLEC, and hence the two rates $r_{\text{ILEC}} = num_{\text{ILEC}}/denom_{\text{ILEC}}$ and $r_{\text{CLEC}} = num_{\text{CLEC}}/denom_{\text{CLEC}}$.

2. Calculate the difference between the two sample rates; if *larger* CLEC rate indicates worse performance, use $DIFF = r_{\text{CLEC}} - r_{\text{ILEC}}$, otherwise take the negative of this.

3. Calculate an estimate of the *standard error for the difference* in the two rates according to the formula

$$\sigma_{\text{DIFF}} = \sqrt{r_{\text{ILEC}}\left[\frac{1}{denom_{\text{CLEC}}} + \frac{1}{denom_{\text{ILEC}}}\right]}$$

4. Compute the test statistic

$$z = \frac{DIFF}{\sigma_{\text{DIFF}}}$$

5. Determine a critical value c so that the type one error is suitably small.

6. Declare the means to be in violation of parity if $z > c$.

Example:

c:	3.58	Critical value for the test

ILEC			CLEC			Test	
num	den	rate	num	den	rate	z	Violation
250	610	0.409836	34	30	1.133333	6.04	YES!

Attachment C

Permutation Analysis Procedural Steps

Permutation analysis is applied to calculate the z-statistic using the following logic:

1. Choose a sufficiently large number T.

2. Pool and mix the CLEC and ILEC data sets

3. Randomly subdivide the pooled data sets into two pools, one the same size as the original CLEC data set (n_{CLEC}) and one reflecting the remaining data points, (which is equal to the size of the original ILEC data set or n_{ILEC}).

4. Compute and store the Z-test score (Z_S) for this sample.

5. Repeat steps 3 and 4 for the remaining T-1 sample pairs to be analyzed. (If the number of possibilities is less than 1 million, include a programmatic check to prevent drawing the same pair of samples more than once).

6. Order the Z_S results computed and stored in step 4 from lowest to highest.

7. Compute the Z-test score for the original two data sets and find its rank in the ordering determined in step 6.

8. Repeat the steps 2-7 ten times and combine the results to determine P = (Summation of ranks in each of the 10 runs divided by 10T)

9. Using a cumulative standard normal distribution table, find the value Z_A such that the probability (or cumulative area under the standard normal curve) is equal to P calculated in step 8.

10. Compare Z_A with the desired critical value as determined from the critical Z table. If $Z_A >$ the designated critical Z-value in the table, then the performance is non-compliant.

Attachment D

Statistical Demonstrations of Non-Parity are Sufficient: Notes on "Competitive Significance"

Some incumbents have proposed that, when comparing the CLEC data set to the ILEC data set for a particular performance measurement result, a lack of parity should not be declared unless both the performance difference is statistically significant and the difference has "competitive or economic significance." This notion is contrary to FCC's interpretation of the terms of the 1996 Act (the Act). The FCC has found that the term "nondiscriminatory" as used in the Act is a more stringent standard than the "unjust and unreasonable discrimination" standard set forth in other provisions of the Communications Act.[40] Thus, the term "nondiscriminatory access" means that: (1) the quality of performance must be equal among all carriers requesting the support, and (2) where technically feasible, the support must be no less in quality and timeliness than that which the incumbent provides to itself.[41]

Some ILECs have also argued that, as the number of data points underlying the computed performance result increases (all other factors held constant),

[40] See FCC Docket No. 96-98, Implementation of the Local Competition Provisions in the Telecommunications Act of 1996, First Report and Order released August 8, 1996, ¶ 217, 859 ("Local Competition Order").

[41] Local Competition Order, ¶315 (access must be provided on terms that are "equal to the terms and conditions under which the incumbent LEC provisions such elements to itself"); Second Order on Reconsideration, Implementation of the Local Competition Provisions of the Telecommunications Act of 1996, CC Docket No. 96-98 (released December 13, 1996) ¶9 (OSS access "must be equal to" the access that the ILEC provides to itself); FCC CC Docket No. 97-137, In the Matter of Ameritech Michigan Pursuant to Section 271 of the Communications Act of 1934, as amended, To Provide In-Region InterLATA Services in Michigan, Memorandum Opinion and Order released August 19, 1997 ("Ameritech Michigan Order"),¶139 ("BOC must provide access to competing carriers that is equal to the level of access that the BOC provides to itself . . . in terms of

smaller differences in means will be statistically significant. This statement is true; nevertheless, as explained in the text, the consequences defined by this plan do not increase with the number of data points. Therefore, the statistical test and z-score have achieved their exact purposes by *identifying unequal performance* and increasing consequences with *severity* of failure. Furthermore, the term "discriminatory" under the Act should not be confused with direct and provable competitive injury. The language of the Act does not permit the incumbent to discriminate against a CLEC by showing that no specific competitive harm was experienced by the CLEC.[42] Moreover, as a theoretical matter, although statistical science can be used to evaluate the impact of different choices of alternative hypothesis in the balancing methodology, there is not much that an appeal to statistical principles can offer in directing specific choices. These specific choices are best left to telephony experts.

These judgements should consider the financial impact (on the CLECs) of violations of various degrees. As a first approximation, the ILEC has data, generated by its routine management procedures, that could be used to calibrate the effect of various violations. The Commission should require the ILEC to produce evidence, relating to its management procedures, that would help the Commission understand what deviations from target performance routinely signal the need for correction.

It is certainly not sufficient to consider only the resulting critical values or error probabilities.

---

quality, accuracy and timeliness"); ¶166 (ILEC "must provide competing carriers access to such OSS function equal to the access that it provides to its retail operations").

[42] Indeed, requiring a CLEC to demonstrate the specific anticompetitive consequences of an ILEC performance failure would effectively render these new protections into mere reiterations of Section II of the Sherman Act. Long experience under antitrust law shows how difficult and protracted such a requirement is in practice.

Attachment E

Mitigation for Potential Impacts of Random Variation is Unnecessary When Type I and Type II Error is Balanced

Random variation is differences in the expected output (or result) of a process that cannot be entirely explained as a result of differences in the inputs to the process. Said another way, running the very same process multiple times using exactly the same key inputs may not (and likely will not) produce exactly the same outcomes. The differences in the outcomes are "explained" as random variation.

There is little debate that the support processes that incumbents utilize to support CLECs tend to be complex and that a variety of factors influence the quantity and quality of the support delivered. As a result, provided the necessary steps have been taken to disaggregate measurement results sufficiently to account for factors correlated with different outcomes, random variation should be accommodated. In doing so, a reasonable balance needs to be struck between (1) protecting the ILEC from consequences that are a result of random variation, and (2) protecting competitors from the adverse effects of discrimination by the ILEC.

As discussed above, the first step in mitigating the effects of random variation is to minimize the risk of making an incorrect decision. In this situation, the two potential incorrect decisions are (1) declaring performance compliant when it is actually discriminatory and (2) declaring performance non-compliant when it is actually within acceptable limits. If these two probabilities are balanced, then, the consequences for "false" failures conceptually offset the consequences for undetected failures. Otherwise stated, the small remedy payment by the ILEC under falsely declared non-compliance is conceptually balanced with the market losses experienced by the CLECs due to falsely declared compliance.

Some regulators have expressed concerns, in light of what they consider to be sizable consequences necessary to motivate compliant ILEC performance and the inability to precisely balance risk, that additional mitigating factors should be instituted. Unfortunately, virtually all the mechanisms discussed are designed to protect the incumbent at the expense of the protecting the competitive process. The following mechanisms have been proposed, but each suffer from serious flaws.

a. Credits for "Better than Required" Performance Permit Gaming

This approach to mitigation is misguided and has the potential to cause extreme harm with little upside potential. In this flawed approach to mitigation, consequences for failed performance could be negated if the incumbent provides "better than required" performance at a different time (or for a different measurement) and thus earns a "credit." For example, the incumbent could deliver bad performance in one area and offset the consequence through performance credits "earned" in a separate but unrelated area or through credits for compliant performance previously (or subsequently) delivered. In all cases, such credits provide incumbents extensive opportunities to "game the system." Credits give ILECs the opportunity to deliver highly variable results that swing between very good and extremely poor performance and still be absolved of any consequence. Likewise, incumbents have the opportunity to temporarily provide compliant performance and then discriminate with impunity. In either case, the CLECs' position in the marketplace compared to the incumbent is harmed. Moreover, because CLECs only learn of "better" performance after the fact (in a performance report), they cannot take practical advantage of such performance. Thus they get no benefit that offsets the real harm they and their customers have actually suffered.

b. Absolute Caps On Liability Are Unwarranted

There is no logical or practical basis to set an absolute limit on any incumbent's liability under any consequences plan, especially for Tier I type consequences. Such consequences are intended to compensate CLECs for actual harm they have sustained as a result of documented poor performance. Thus, there should never be a limit on this type of consequence. Moreover, to the extent that Tier II consequences become especially large, it may be appropriate to establish a procedural cap to provide an opportunity to assess whether the calculated consequence for an incumbent's market-affecting behavior should be limited.

Attachment F

Addressing Measurement Overlap And Correlation

Measurement overlap occurs when one or more measurements effectively measure the same performance. If two measurements overlap, then consequences should attach to only one of them. Note, however, a measurement addressing timeliness and a measurement addressing quality for the same area of performance do not overlap. Also, it should be noted that, given the care taken in defining measurements in LCUG SQM Version 7.0, there are no obvious areas of significant measurement overlap

Measurement correlation is different from measurement overlap. Measurement correlation occurs when one or more measurement results move at the same time. The direction of movement need not be the same. That is, one may improve (e.g., quality) while another deteriorates (e.g., timeliness). As such, measurement correlation does not automatically argue for adjustment to the measurements eligible for consequences. Indeed, an incumbent that is intentionally and pervasively discriminating would be capable of showing a high degree of correlation among all measurement results both within and across months – all results would be deteriorating.

If there are reasons to believe that measurements are somewhat overlapping and correlation is suspected, the solution is not to immediately eliminate one or both measurements. Rather the potentially superior approach is to create "families" for the purpose of applying consequences. Each measurement "family" would be eligible for only a single consequence. Whether and to what degree a family is eligible for a consequence would be determined by the worst performing individual measurement result within the family for the month under consideration. Thus, use of measurement families eliminates the possibility of consequence "double jeopardy"[43] without making any advance value judgement regarding the usefulness of individual measurements.

Use of measurement families has the potential for significant harm for an otherwise effective consequence plan due because: (1) inappropriate grouping can mask areas of discrimination by placing non-overlapped measurements in the same family; and, (2) by reducing eligible measurements, without adjusting the per measurement consequence, the overall plan incentives are diminished. As a result, establishment of measurement families must be approached with extreme caution and sparingly used. At least the following conditions must be imposed.

(1) measurements that address separate support functionality may not be placed in the same family;

(2) measurements that address different modes of market entry may not be placed in the same family;

(3) measurement families may not be used as a means to avoid disaggregation detail;

(4) measurements that address (a) timeliness, (b) accuracy, and (c) completeness may not be placed within the same family;

(5) measurement families, to the extent used, must be identical across all CLECs;

(6) even if correlation can be demonstrated, measurement families must not be used to combine otherwise independent measurements of a deficient process; and,

[43] If the measurements in the family are truly overlapping and correlated they point to the same conclusion (incidents of failure and severity). Measurement families thus treat the incumbent preferentially: either the measurements are effectively the same and only one consequence applies or they were inappropriately grouped and the incumbent avoids one or more consequences that should have been incurred.

(7) establishment of measurement families must not reduce the maximum consequence payable by more than 10% without an offsetting increase in the basic, intermediate, and severe consequence payable per failed measurement.

To the extent new measurement families are proposed or a proposal is set forth to eliminate or modify and existing family, the advocate of the change should bear the burden of demonstrating compliance with the above minimum requirements. The consideration should be in a public forum where all interested parties participate, and in the event of a disagreement, the Commission should decide based upon the record established. Prospective changes of measurement families should not affect any prior determinations regarding consequences.

No proposal to establish measurement families should be considered until the consequence plan has been operational and produced at least six months of independently verified data.

Attachment G

Balancing the Type I and Type II Error Probabilities of the Modified Z Test Statistic

This appendix describes the methodology for balancing the error probabilities when the Modified Z statistic is used for performance measure parity testing. There are four key elements of the statistical testing process:

1. the null hypothesis, H_0, that parity exists between ILEC and CLEC services
2. the alternative hypothesis, H_a, that the ILEC is giving better service to its own customers
3. the Modified Z test statistic, Z, and
4. a critical value, c

The decision rule[44] is

- If $Z < c$ then accept H_a.
- If $Z \geq c$ then accept H_0.

There are two types of error possible when using such a decision rule:

Type I Error: Deciding favoritism exists (accept H_a) when there is, in fact, no favoritism (H_0 is true).

Type II Error: Deciding parity exists (accept H_0) when there is, in fact, favoritism (H_a is true).

The probabilities of the two types of error are:

Type I Error: $\alpha = P(Z < c \mid H_0)$.

Type II Error: $\beta = P(Z \geq c \mid H_a)$.

In what follows, we show how to find a balancing critical value, c_B, so that $\alpha = \beta$.

General Methodology

The general form of the test statistic that is being used is

$$z_0 = \frac{\hat{T} - E(\hat{T}|H_0)}{SE(\hat{T}|H_0)}, \tag{1}$$

where

$\hat{T}$ is an estimator that is (approximately) normally distributed,

$E(\hat{T} \mid H_0)$ is the expected value (mean) of $\hat{T}$ under the null hypothesis, and

[44] This decision rule assumes that the smaller a performance measure is, the better the service. If the opposite is true, then the decision rule should be reversed by using –Z in place of Z.

$SE(\hat{T} \mid H_0)$ is the standard error of $\hat{T}$ under the null hypothesis.

Thus, under the null hypothesis, z_0 follows a standard normal distribution. However, this is not true under the alternative hypothesis. In this case,

$$z_a = \frac{\hat{T} - E(\hat{T} | H_a)}{SE(\hat{T} | H_a)}$$

has (approximately) a standard normal distribution. Here

$E(\hat{T} \mid H_a)$ is the expected value (mean) of $\hat{T}$ under the alternative hypothesis, and

$SE(\hat{T} \mid H_a)$ is the standard error of $\hat{T}$ under the alternative hypothesis.

Notice that

$$\begin{aligned} \beta &= P(z_0 > c \mid H_a) \\ &= P\left(z_a > \frac{cSE(\hat{T} \mid H_0) + E(\hat{T} \mid H_0) - E(\hat{T} \mid H_a)}{SE(\hat{T} \mid H_a)} \right) \end{aligned} \tag{2}$$

and recall that for a standard normal random variable z and a constant b, $P(z < b) = P(z > -b)$. Thus,

$$\alpha = P(z_0 < c) = P(z_0 > -c) \tag{3}$$

Since we want $\alpha = \beta$, the right hand sides of (2) and (3) represent the same area under the standard normal density. Therefore, it must be the case that

$$-c = \frac{cSE(\hat{T} \mid H_0) + E(\hat{T} \mid H_0) - E(\hat{T} \mid H_a)}{SE(\hat{T} \mid H_a)}.$$

Solving this for c give the general formula for a balancing critical value:

$$c_B = \frac{E(\hat{T} \mid H_a) - E(\hat{T} \mid H_0)}{SE(\hat{T} \mid H_a) + SE(\hat{T} \mid H_0)} \tag{4}$$

The Balancing Critical Value of the Modified Z for a Mean Measure

The modified Z statistic Z for a mean measure is given by

$$Z = \frac{\hat{T}}{s_1\sqrt{1/n_1 + 1/n_2}}$$

where $\hat{T} = \overline{X}_1 - \overline{X}_2$ and subscripts 1 and 2 refer to ILEC and CLEC quantities, respectively.

One possible set of hypotheses, that take into account the assumption that transaction are identically distributed within LECs, is:

H_0: $\mu_1 = \mu_2$, $\sigma_1^2 = \sigma_2^2$

H_a: $\mu_2 = \mu_1 + \delta\cdot\sigma_1$, $\sigma_2^2 = \lambda\cdot\sigma_1^2$ $\qquad$ $\delta > 0$ and $\lambda \geq 1$.

Assuming that n_1 is large enough so that s_1 adequately approximates σ_1, we have

$$E(\hat{T} \mid H_0) = 0$$

$$SE(\hat{T} \mid H_0) = \sigma_1\sqrt{1/n_1 + 1/n_2}$$

$$E(\hat{T} \mid H_a) = -\delta\sigma_1$$

$$SE(\hat{T} \mid H_a) = \sigma_1\sqrt{1/n_1 + \lambda/n_2}$$

Substituting these values in equation (4) gives

$$c_B = \frac{-\delta}{\sqrt{1/n_1 + 1/n_2} + \sqrt{1/n_1 + \lambda/n_2}}$$

$$= \frac{-\delta\sqrt{n_1 n_2}}{\sqrt{n_1 + n_2} + \sqrt{\lambda n_1 + n_2}}$$

The preceding equations have indexed the alternative hypothesis by two parameters, λ and δ. While statistical science can be used to evaluate the impact of different choices of these parameters, there is not much that an appeal to statistical principles can offer in directing specific choices. Specific choices are best left to telephony experts. Still, it is possible to comment on some aspects of these choices:

Parameter Choice for λ. The parameter λ indexes an alternative to the null hypothesis that arises because there might be greater unpredictability or variability in the delivery of service to a CLEC customer over that which would be achieved for an otherwise comparable ILEC customer. Typically, there is little basis for choosing a value of λ other than 1, in which case the formula for c_B simplifies to

$$c_B = \frac{-\delta\sqrt{n_1 n_2}}{2\sqrt{n_1 + n_2}}$$

Parameter Choice for δ. The parameter δ is much more important in the choice of the balancing point than was true for λ because it directly indexes the difference in average service. The Joint CLEC's Performance Incentive Plan proposes that δ should be no greater than 0.25.

The Balancing Critical Value of the Modified Z for a Proportion Measure

Specification of a balancing critical value for a proportion measure is more complex than for mean measures because c_B depends directly on both the assumed ILEC and CLEC proportions under H_a not just through a single parameter like δ.

The modified Z statistic for a proportion measure is given by

$$Z = \frac{\hat{T}}{\sqrt{\hat{p}_{ILEC}(1-\hat{p}_{ILEC})}\sqrt{1/n_1 + 1/n_2}}$$

where $\hat{T} = \hat{p}_{ILEC} - \hat{p}_{CLEC}$, and where n_1 and n_2 are the ILEC and CLEC sample sizes, respectively.

The null and alternative hypotheses are specified fully in terms of the true proportions p_{ILEC} and p_{CLEC} as follows:

H_0: $p_{ILEC} = p_{CLEC} = p_1$

H_a: $p_{ILEC} = p_1$, $p_{CLEC} = p_2 > p_1$

Assuming that n_1 is large enough so that $\hat{p}_{ILEC}(1-\hat{p}_{ILEC})$ adequately approximates $p_{ILEC}(1-p_{ILEC})$, then Z satisfies (1) and we have

$$E(\hat{T} \mid H_0) = 0$$

$$SE(\hat{T} \mid H_0) = \sqrt{p_1(1-p_1)}\sqrt{1/n_1 + 1/n_2}$$

$$E(\hat{T} \mid H_a) = p_1 - p_2$$

$$SE(\hat{T} \mid H_a) = \sqrt{p_1(1-p_1)/n_1 + p_2(1-p_2)/n_2}$$

Substituting these values in equation (4) gives

$$c_B = \frac{-(p_2 - p_1)}{\sqrt{p_1(1-p_1)/n_1 + p_2(1-p_2)/n_2} + \sqrt{p_1(1-p_1)}\sqrt{1/n_1 + 1/n_2}}.$$

A convenient way to specify the alternative hypothesis is through the "odds ratio" for p_2 and p, specifically

$$\varphi = \left(\frac{p_2}{p_1}\right)\left(\frac{1-p_1}{1-p_2}\right)$$

so that

$$p_2 = \left(\frac{\varphi p_1}{1+(\varphi-1)p_1}\right).$$

The Balancing Critical Value of the Modified Z for a Rate Measure

A rate is a ratio of two counts *num/denom*—e.g., $r_{ILEC} = num_{ILEC}/denom_{ILEC}$--where the *denom* count is assumed known but the *num* count is subject to sampling variability. Similarly to proportions, the balancing critical value c_B depends directly on the assumed ILEC and CLEC rates under H_a as well as the ILEC and CLEC denominators.

The modified Z statistic for a rate measure is given by

$$Z = \frac{\hat{T}}{\sqrt{\hat{r}_{ILEC}(1/denom_{CLEC} + 1/denom_{ILEC})}}$$

where $\hat{T} = \hat{r}_{ILEC} - \hat{r}_{CLEC}$.

The null and alternative hypotheses are specified fully in terms of the true proportions r_{ILEC} and r_{CLEC} as follows:

H_0: $r_{ILEC} = r_{CLEC} = r_1$

H_a: $r_{ILEC} = r_1$, $r_{CLEC} = r_2 > r_1$

Assuming that $denom_{ILEC}$ is large enough so that $\hat{r}_{ILEC}$ adequately approximates r_{ILEC}, then Z satisfies (1) and we have

$$E(\hat{T} \mid H_0) = 0$$

$$SE(\hat{T} \mid H_0) = \sqrt{r_{ILEC}(1/denom_{CLEC} + 1/denom_{ILEC})}$$

$$E(\hat{T} \mid H_a) = r_1 - r_2$$

$$SE(\hat{T} \mid H_a) = \sqrt{r_{CLEC}/denom_{CLEC} + r_{ILEC}/denom_{ILEC})}$$

Substituting these values in equation (4) gives

$$c_B = \frac{-(r_2 - r_1)}{\sqrt{r_{CLEC}/denom_{CLEC} + r_{ILEC}/denom_{ILEC})} + \sqrt{r_{ILEC}(1/denom_{CLEC} + 1/denom_{ILEC})}}$$

A convenient way to specify the alternative hypothesis is by

$$r_2 = \varepsilon r_1 .$$

CLEC Proposed Disaggregation
(Process Level)

Disaggregation
A. Pre-Order OSS Responsiveness 1. Feature Function Availability/Service Availability 2. Facility Availability Qualification of Loops for Advanced Digital Services 3. Street Address Validation 4. Appointment Scheduling 5. Customer Service Records 6. Telephone Number 7. Rejected or Failed Queries (regardless of type)
B. Maintenance & Repair OSS Responsiveness 1. Create (or confirm logging of) a Maintenance Request 2. Obtain Status 3. Obtain Test Results 4. Cancel Request 5. Rejected of Failed Queries (regardless of type) 6. Clearance Notification 7. Closure Notification
C. Collocation 1. Physical Caged 2. Shared Caged 3. Cageless 4. Adjacent On-Site 5. Adjacent Off-Site 6. Augment to Physical 7. Virtual 8. Augment to Virtual
D. Multi-Functional Disaggregation 1. Interface type—for preordering, ordering, billing and maintenance and repair OSS 2. Dispatch and non-dispatch—for provisioning and maintenance measures 3. Volume—for ordering, provisioning, and maintenance measures (a) 1-5 lines, (b) 6-14 lines, and (c) 15+ lines 4. Geographic --All measures should be disaggregated to a state level, if the data is available. Additionally, provisioning and maintenance measures should be disaggregated to the MSA level 5. By CLEC, BST, and all BST affiliates for all measures 6. Center—for OS/DA, ordering & maintenance service center measures
E. Service Order Activities 1. New Service Installations 2. Service Migrations Without Changes 3. Service Migrations With Changes 4. Local Number Porting 5. Inside Move 6. Outside Move 7. Records Change 8. Feature Changes 9. Service Disconnects 10. Translation Disconnects 11. Standalone Directory Listing (DL) 12. Standalone Directory Assistance (DA) Listing

Disaggregation
13. Standalone DL & DA Activity
F. Billing 1. Record Type (resale, interconnection, UNE)

Disaggregation, Analogs and Benchmarks		
G. Product Level Disaggregation for (Ordering, Provisioning, and Maintenance & Repair)	**Benchmark– 95% within x Days unless otherwise noted (resale) for Order Completion Interval**	**Retail analog for other provisioning and maintenance and repair measures**
1. Resold Residence POTS	1. Retail Analog	1. Retail Analog
2. Resold Business POTS	2. Retail Analog	2. Retail Analog
3. Resold BRI ISDN	3. Retail Analog	3. Retail Analog
4. Resold PRI ISDN	4. Retail Analog	4. Retail Analog
5. Resold Centrex/Centrex-like	5. Retail Analog	5. Retail Analog
6. Resold Analog PBX trunks	6. Retail Analog	6. Retail Analog
7. Resold DID Trunks	7. Retail Analog	7. Retail Analog
8. Resold Voice-Grade Private Line	8. Retail Analog	8. Retail Analog
9. Resold DS1 Services	9. Retail Analog	9. Retail Analog
10. Resold DS3 Services	10. Retail Analog	10. Retail Analog
11. Resold >DS3 Services	11. Retail Analog	11. Retail Analog
12. Other Resold Services	12. Retail Analog	12. Retail Analog
13. UNE Platform	13. Retail POTS	13. Retail POTS
14. UNE Channelized DS1 (DS1 loop + multiplexing)	14. 3, 7, and 10 days, for a ,b, and c, volumes respectively	14. DS1
15. Unbundled 8 dB Analog Loops	15. Same as above	15. Retail POTS
16. Unbundled 2-wire Digital Loops	16. Same as above	16. Retail POTS
17. Unbundled 4-wire Digital Loops	17. Same as above	17. Retail POTS
18. Unbundled ADSL Loops	18. Same as above	18. DS1
19. Unbundled HDSL Loops	19. Same as above	19. DS1
20. Unbundled xDSL Loops	20. Same as above	20. DS1
21. Other Unbundled Loops	21. Same as above	21. DS1
22. UNE Analog Switch Port (line side)	22. 2 days	22. POTS
23. UNE BRI Capable Switch Port (line side)	23. 3 days	23. ISDN
24. UNE DS1 Switch Port (line side)	24. 5 days	24. DS1
25. UNE PRI Switch Port (trunk side)	25. 5 days	25. ISDN
26. UNE DID-capable Switch Port (trunk side)	26. 5 days	26.
27. UNE Message Trunk Port	27. 5 days	27. DS1
28. UNE Dedicated DS0 Transport	28. 3, 7, and 10 days, for a ,b, and c, volumes respectively	28. DS1
29. UNE Dedicated DS1 Transport	29. Same as above	29. DS1
30. UNE Dedicated DS3 Transport	30. Same as above	30. DS3
31. Interconnect Trunks (DS0s, DS1s and DS3s,)	31. ILEC Trunks	31. ILEC Trunks
32. Two-Way Trunking, Inbound Augments, separately)	32. ILEC Trunks	32. ILEC Trunks

Disaggregation, Analogs and Benchmarks		
G. Product Level Disaggregation for (Ordering, Provisioning, and Maintenance & Repair)	**Benchmark– 95% within x Days unless otherwise noted (resale) for <u>Order Completion Interval</u>**	**Retail analog for other provisioning and maintenance and repair measures**
33. ILNP	33. 3, 7, and 10 days, for a ,b, and c, volumes respectively	33. Retail POTS
34. PNP or LNP	34. Same as above	**34.** Retail POTS
35. Line-sharing/High Frequency Spectrum UNE	35. 3, 5 and 7 days for a, b and c, volumes	**35.** Retail POTS
36. Sub-loop unbundling, e.g. network terminating wire	36. 5, 7, 10 days for a, b, and c, volumes	**36.** Retail POTS
37. Loop Modification/Loop Conditioning	37. 5, 7, 10 days for a, b,and c volumes.	**37.** Retail POTS

Measure Concentrated CLECs and the Maximum Balancing Critical Value

CLECs are often highly specialized in the services they provide and how they are provided. Some CLECs only resale local service, some only use UNE-Platform, and some only provide DSL service. Others, of course, may provide a broad range of services. For highly specialized CLECs, the number of performance measurements populated in a given month will be relatively small (as compared to those that provide a wider range of services) and the sample sizes in those metrics relatively large. It is important that the Performance Incentive Plan (PIP) not discriminate against CLECs based on business decisions such as what services to provide and how they are provided. A balancing critical value approach that fails to cap the balancing critical value discriminates against the "large sample" CLEC and its customers

Recall that the balancing critical value can be *approximated* by the formula:

$$C_B = \frac{-\delta}{2\sqrt{1/n_c}} = -\delta \cdot 0.5\sqrt{n_c} \qquad (1)$$

where n_c is the CLEC sample size. Equation (1) shows that as the CLEC sample size increases, so does the balancing critical value. Thus, the larger is the CLEC sample size, the more difficult it is to detect a means difference because the significance level of means test is increasing with sample size.

While means difference tests are traditionally conducted at a significance level no smaller than 0.01, the balancing approach uses a significances level of 0.0000000000000035 (C_B = -7.91) for a CLEC sample size of 1,000 (which, in fact, is not that large). Obviously, it is considerably more difficult to detect a means difference at a significance level of 0.0000000000000035 than 0.01. Further, if the expected consequence of Type I error (payment for a false positive) is non-existent at a significance level of 0.001, there is no reason to raise the bar to a significance level of 0.0000000000000035.